

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:50:23 ; Search time 394.5 Seconds
(without alignments)
12915.082 Million cell updates/sec

Title: US-09-787-016A-1

Sequence: 1 cccggtggtccgcgcgcac.....actcttaagatcatatccg 2610

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 80%

Maximum Match 100%
Listing first 1000 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgcn2_1/USPTO.spool.p/US09787016/runat_20042004_154658_17003/app.query.fasta_1.2759
-DB=Pending Patents AA Main -OPMT=fastan -SUFFIX=pct.rapm -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=1000 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=80 -ALIGN=100 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000
-USER=US09787016 @CGN 1.1 444 @runat_20042004_154658_17003 -NCPU=6 -ICPU=3
-NO MMAP -LARGEJQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MAPN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Pending_Patents_AA_Main:*

1: /cgcn2_6/prodata/2/paa/PCTUS_COMB.pep:*
2: /cgcn2_6/prodata/2/paa/US06_COMB.pep:*
3: /cgcn2_6/prodata/2/paa/US07_COMB.pep:*
4: /cgcn2_6/prodata/2/paa/US080_COMB.pep:*
5: /cgcn2_6/prodata/2/paa/US081_COMB.pep:*
6: /cgcn2_6/prodata/2/paa/US082_COMB.pep:*
7: /cgcn2_6/prodata/2/paa/US083_COMB.pep:*
8: /cgcn2_6/prodata/2/paa/US084_COMB.pep:*
9: /cgcn2_6/prodata/2/paa/US085_COMB.pep:*
10: /cgcn2_6/prodata/2/paa/US086_COMB.pep:*
11: /cgcn2_6/prodata/2/paa/US087_COMB.pep:*
12: /cgcn2_6/prodata/2/paa/US088_COMB.pep:*
13: /cgcn2_6/prodata/2/paa/US089_COMB.pep:*
14: /cgcn2_6/prodata/2/paa/US090_COMB.pep:*
15: /cgcn2_6/prodata/2/paa/US091_COMB.pep:*
16: /cgcn2_6/prodata/2/paa/US092_COMB.pep:*
17: /cgcn2_6/prodata/2/paa/US093_COMB.pep:*
18: /cgcn2_6/prodata/2/paa/US094_COMB.pep:*
19: /cgcn2_6/prodata/2/paa/US095_COMB.pep:*
20: /cgcn2_6/prodata/2/paa/US096_COMB.pep:*
21: /cgcn2_6/prodata/2/paa/US097A_COMB.pep:*
22: /cgcn2_6/prodata/2/paa/US097B_COMB.pep:*
23: /cgcn2_6/prodata/2/paa/US098_COMB.pep:*
24: /cgcn2_6/prodata/2/paa/US099A_COMB.pep:*
25: /cgcn2_6/prodata/2/paa/US099B_COMB.pep:*
26: /cgcn2_6/prodata/2/paa/US100_COMB.pep:*
27: /cgcn2_6/prodata/2/paa/US101_COMB.pep:*

28: /cgcn2_6/prodata/2/paa/US102_COMB.pep:*
29: /cgcn2_6/prodata/2/paa/US103_COMB.pep:*
30: /cgcn2_6/prodata/2/paa/US104_COMB.pep:*
31: /cgcn2_6/prodata/2/paa/US106_COMB.pep:*
32: /cgcn2_6/prodata/2/paa/US107_COMB.pep:*
33: /cgcn2_6/prodata/2/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
No matches found					

Search completed: April 20, 2004, 21:12:13
Job time : 394.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:51:33 ; Search time 19.5 Seconds
(without alignments)
7386.308 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 ctcggtgagccgcgcacac.....actcttaagatcatatccg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 196679 seqs, 27592530 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cg2_1/USPTO.spool.p/US09787016/runat.20042004.154659.17029/app.query.fasta_1.2759
-DB=Pending Patents AA New -QFMT=faстан -SUFFIX=pct.rapn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=1000 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=80 -ALIGN=100 -MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=2000000000
-USER=US09787016 @CGN 1.1 62 @runat.20042004.154659.17029 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Pending_Patents_AA_New:*
1: /cg2_6/prodata/1/paa/PCT_NEW_COMB.pep:*
2: /cg2_6/prodata/1/paa/US06_NEW_COMB.pep:*
3: /cg2_6/prodata/1/paa/US07_NEW_COMB.pep:*
4: /cg2_6/prodata/1/paa/US08_NEW_COMB.pep:*
5: /cg2_6/prodata/1/paa/US09_NEW_COMB.pep:*
6: /cg2_6/prodata/1/paa/US10_NEW_COMB.pep:*
7: /cg2_6/prodata/1/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

No matches found					

Search completed: April 20, 2004, 21:13:05
Job time : 19.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_nzp model

Run on: April 20, 2004, 20:45:13 ; Search time 43 Seconds
(without alignments)
11677.204 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 ctcgctgcgcgcgcgcacac.....acctctaagatcatatcccg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:

-MODEL=frame+ .n2p.model -DEV=xlp
-O=/cgn2.1/USPTO.spool.p/US09787016/runat.20042004.154657.16941/app.query.fasta_1.2759
-DB=PIR_78 -OPMT=fastran -SUFFIX=pct.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdd -LIST=1000
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=80 -ALIGN=100 -MODE=LOCAL
-OUTFMT=pcio -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US09787016 @CGN 1.1 66 @runat.20042004.154657.16941 -NCPU=6 -ICPU=3
-NO MMAP -LARGEJQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
No matches found					

Search completed: April 20, 2004, 20:57:39
Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:41:28 ; Search time 28 Seconds
(without alignments)
9707.365 Million cell updates/sec

Title: US-09-787-016A-1

Perfect score: 4689
Sequence: 1 ctcgtrgscgcgcgcacac.....accttaagatcatatcccg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%

Maximum Match 100%
Listing first 1000 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xip
-O=/cgn2.1/USPTO.spool.p/US09787016/runat.20042004.154656.16910/app.query.fasta_1.2759
-DB=SwissProt.42 -QFMT=fastcan -SUPPLY=pct_rap -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=1000 -DOCCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=80 -ALIGN=100
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US09787016 @CGN.1.1.28 @runat.20042004.154656.16910 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEIOBERRY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	ID	Description
No.						

No matches found						

Search completed: April 20, 2004, 20:51:28
Job time : 28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:44:43 ; Search time 129.5 Seconds
(without alignments)
12718.185 Million cell updates/sec

No matches found

Search completed: April 20, 2004, 20:56:00
Job time : 129.5 secs

Title: US-09-787-016a-1
Perfect score: 4689
Sequence: 1 cctcggtgcgcgcgcgcac.....accttaagatcatatccg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool.p/US09787016/runat.20042004.154657.16922/app.query.fasta_1.2759
-DB=SPTREMBL.25 -OPMT=fastan -SUFFIX=pct.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -INITs=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human40.cdi
-LIST=1000 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=80 -ALIGN=100
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016 @CGN 1.1 142 @runat.20042004.154657.16922 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL.25.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvivirus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
No.							

Gencore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:46:13 ; Search time 29.5 Seconds
(without alignments)
9135.172 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 ctcggtgcccgcgcgccac.....actcttaagatcatatcccg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 80%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cg2_1/USPTO.spool_p/US09787016/runat_20042004_154658_16977/app_query.fasta_1.2759
-DB=Issued_Patents_AA -OPT=fastan -SUFFIX=pct.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -INITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=1000 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=80 -ALIGN=100
-MODE=LOCAL -OUTFMT=pct -NORMext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09787016.@CGN_1_1_29.@runat_20042004_154658_16977 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=130 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA.*
1: /cg2_6/prodata/2/1aa/5A.COMB.pep.*
2: /cg2_6/prodata/2/1aa/5B.COMB.pep.*
3: /cg2_6/prodata/2/1aa/5A.COMB.pep.*
4: /cg2_6/prodata/2/1aa/5B.COMB.pep.*
5: /cg2_6/prodata/2/1aa/PTUS.COMB.pep.*
6: /cg2_6/prodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
No matches found				

Search completed: April 20, 2004, 20:58:51
Job time : 29.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:56:04 ; Search time 115 Seconds
(without alignments)
12513.164 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 ctcggtgcccgcgcgccac.....actcttaagatcatatccg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1124875 seqs, 275673149 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 80%

Maximum Match 100%
Listing first 1000 summaries

Command line parameters:

-MODEL=fixme+ n2p.model -DEV=x1p
-Q=/cg2_1/USPTO.spool.p/US09787016/runat.20042004.154659.17051/app.query.fasta_1.2759
-DB=Published Applications AA -QFMT=fastan -SUPFIX=pct.rapb -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 -UNIT=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human0.cdi -LIST=1000 -DOCALLGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=80 -ALIGN=100 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000
-USER=US09787016 @CGN 1.1 17 runat.20042004.154659.17051 -NCPU=6 -ICPU=3
-NO MMAP -LARGEOUTERY -NEG_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:*

1: /cg2_6/prodata/1/pubppa/US07_PUBCOMB.pep:*
2: /cg2_6/prodata/1/pubppa/PCT_NEW_PUB.pep:*
3: /cg2_6/prodata/1/pubppa/US06_NEW_PUB.pep:*
4: /cg2_6/prodata/1/pubppa/US06_PUBCOMB.pep:*
5: /cg2_6/prodata/1/pubppa/US07_NEW_PUB.pep:*
6: /cg2_6/prodata/1/pubppa/PCTUS_PUBCOMB.pep:*
7: /cg2_6/prodata/1/pubppa/US08_NEW_PUB.pep:*
8: /cg2_6/prodata/1/pubppa/US08_PUBCOMB.pep:*
9: /cg2_6/prodata/1/pubppa/US09A_PUBCOMB.pep:*
10: /cg2_6/prodata/1/pubppa/US09B_PUBCOMB.pep:*
11: /cg2_6/prodata/1/pubppa/US09C_PUBCOMB.pep:*
12: /cg2_6/prodata/1/pubppa/US09_NEW_PUB.pep:*
13: /cg2_6/prodata/1/pubppa/US10A_PUBCOMB.pep:*
14: /cg2_6/prodata/1/pubppa/US10B_PUBCOMB.pep:*
15: /cg2_6/prodata/1/pubppa/US10C_PUBCOMB.pep:*
16: /cg2_6/prodata/1/pubppa/US10_NEW_PUB.pep:*
17: /cg2_6/prodata/1/pubppa/US60_NEW_PUB.pep:*
18: /cg2_6/prodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysts of the total score distribution.

No matches found

Search completed: April 20, 2004, 21:17:07
Job time : 116 secs

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
---------------	----------------	-------	--------	----	-------------

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 20, 2004, 20:40:53 ; Search time 119.5 Seconds
(without alignments)
12342.242 Million cell updates/sec

Title: us-09-787-016a-1

Perfect score: 4689
Sequence: 1 ctcggtgcgcgcgcgcac.....actcttaagatcatatccg 2610

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 80%

Maximum Match 100%
Listing first 1000 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cg2.1/USPTO.spool.p/US09787016/funat.20042004.154656.16900/app.query.fasta_1.2759
-DB=A.Geneseq.29Jan04 -QFMT=faetan -SUPFIX=pct.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human4.cdi
-LIST=1000 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=80 -ALIGN=100
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09787016 @CGN 1 1 91 @funat.20042004.154656.16900 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq.29Jan04.*

1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
No matches found					

Search completed: April 20, 2004, 20:50:20
Job time : 119.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2004, 21:12:19 ; Search time 45 seconds
(without alignments)
4305.071 Million cell updates/sec

Title: US-09-787-016a-4
Perfect score: 3228
Sequence: 1 MDCKHLSNEAPRAIKFTS.....LKSCVGLMELATSYSPRPW 614

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*
15: sp_virins:*
16: sp_Bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	323.5	10.0	2016	5 Q9VG78	Q9VG78 drosophila
2	228	7.1	2649	5 Q9W0T2	Q9W0T2 drosophila
3	228	7.1	2669	5 Q9W0T1	Q9W0T1 drosophila
4	225	7.0	2662	5 Q95VB8	Q95VB8 drosophila
5	210.5	6.5	5442	5 Q818F0	Q818F0 euploies cr
6	209	6.5	663	5 Q9W352	Q9W352 drosophila
7	200.5	6.2	3146	5 Q9VUB5	Q9VUB5 drosophila
8	200	6.2	366	5 Q8T3Y1	Q8T3Y1 drosophila
9	198.5	6.1	424	3 Q74508	Q74508 schizosacch
10	196.5	6.1	645	11 Q8VDN7	Q8VDN7 mus musculu
11	191.5	5.9	2492	6 Q7YQM3	Q7YQM3 pongo pygma
12	188.5	5.8	569	4 Q9GCT5	Q9GCT5 homo sapien
13	188.5	5.8	670	4 Q9BM08	Q9BM08 homo sapien
14	188.5	5.8	671	4 Q724V5	Q724V5 homo sapien
15	188.5	5.8	2764	4 Q727D6	Q727D6 homo sapien
16	187.5	5.8	1893	11 Q8CJ14	Q8CJ14 rattus norv

17	187.5	5.8	2781	4 Q9UG2	Q9UG2 homo sapien
18	183.5	5.7	2492	6 Q7YQM4	Q7YQM4 pan troglod
19	179	5.5	385	5 Q17909	Q17909 caenorhabdi
20	179	5.5	702	4 Q92541	Q92541 homo sapien
21	178.5	5.5	451	5 Q9NLC1	Q9NLC1 caenorhabdi
22	177	5.5	443	11 Q8C969	Q8C969 mus musculu
23	177	5.5	473	11 Q8C980	Q8C980 mus musculu
24	177	5.5	1110	13 Q91255	Q91255 petromyzon
25	170.5	5.3	405	5 Q45410	Q45410 caenorhabdi
26	170.5	5.3	878	13 Q8AVW4	Q8AVW4 xenopus lae
27	169.5	5.3	781	4 Q8NAP7	Q8NAP7 homo sapien
28	169.5	5.3	17352	5 Q95YM2	Q95YM2 procambaris
29	169	5.2	5412	5 Q9W596	Q9W596 drosophila
30	168	5.2	3279	5 Q9N4B9	Q9N4B9 caenorhabdi
31	167	5.2	5327	5 Q76891	Q76891 drosophila
32	166	5.1	688	2 Q9X4J3	Q9X4J3 ehrlichia c
33	165.5	5.1	563	13 Q7SZK6	Q7SZK6 brachydanto
34	165.5	5.1	563	13 Q7SZB4	Q7SZB4 brachydanto
35	165.5	5.1	2400	4 Q81WP2	Q81WP2 homo sapien
36	165.5	5.1	2416	4 Q81WP1	Q81WP1 homo sapien
37	165.5	5.1	2432	4 Q81WP0	Q81WP0 homo sapien
38	165.5	5.1	2448	4 Q81WN9	Q81WN9 homo sapien
39	165.5	5.1	2464	4 Q81WN8	Q81WN8 homo sapien
40	165.5	5.1	2480	4 Q81WN7	Q81WN7 homo sapien
41	165	5.1	572	6 Q8MKF9	Q8MKF9 canis fam11
42	165	5.1	4498	13 Q93291	Q93291 fugu rubrip
43	165	5.1	8081	5 Q72120	Q72120 caenorhabdi
44	164.5	5.1	510	5 Q45407	Q45407 caenorhabdi
45	164.5	5.1	1444	5 Q9VTN2	Q9VTN2 drosophila

ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	2016 AA.
Q9VG78	Q9VG78		
AC	Q9VG78;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DB	CG6525 protein.		
GN	SGP OR CG6525.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OK	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins K.A., Galie R.F.,		
RA	George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Abdill J.F., Agbayani A., An H.-U., Andrews-Fiankoch C., Baldwin D.,		
RA	Baillet R.M., Baas A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,		
RA	Borova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,		
RA	Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Fabros B., Delcher A., Deng Z., Mayh A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,		
RA	Foster C., Gabrielian A.B., Gary N.S., Gelbart W.M., Glaser K.,		
RA	Gooley A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,		
RA	Hoslin D., Houston K.A., Howland T.U., Wei M.-H., Ibegwam C.,		
RA	Jalili W., Janssen F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		

RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,
RA Paciel J., Parasas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Piltman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT *Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Miera S., Creaby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker S.B., Prochink S.B., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield B.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT *Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celinker S.B., Glibbs R.A., Rubin G.M., Venter J.C.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB003467; AAF47361.2; -
DR FlyBase: FBgn0000541; E(bx).
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0005489; F:electron transporter activity; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR000637; AT hook.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR000345; Cytochrome B5.
DR InterPro: IPR004022; DGT-dom.
DR InterPro: IPR006209; EGF-like.
DR InterPro: IPR001965; ZnF-PHD.
DR Pfam: PF02178; AT hook; 1.
DR Pfam: PF00439; bromodomain; 1.
DR Pfam: PF02791; DDT; 1.
DR Pfam: PF00628; PHD; 3.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00249; PHD; 3.
DR SMART: SM00249; BROMO; 1.
DR PROSITE: PS00633; BROMODOMAIN_1; 1.
DR PROSITE: PS00114; BROMODOMAIN_2; 1.
DR PROSITE: PS00190; CYTOCHROME_C_1.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01359; ZF-PHD_1; 1.
DR PROSITE: PS00016; ZF-PHD_2; 2.
DR PROSITE: PS00016; ZF-PHD_2; 2.
SQ SEQUENCE 2649 AA; 298507 MW; C0D7F7E015EA5403 CRC64;

Query Match 7.1% Score 228; DB 5; Length 2649;

Best Local Similarity 22.1%; Pred. No. 9e-07; Matches 14;

Matches 77; Conservative 52; Mismatches 98; Indels 122; Gaps 14;

93 DSSEPTSTVTV-ETASEGVSSESRISGVSLSL-----GKHPASSEKAGGE 145

2197 EHDPTNLACIDISTDLEKONKESFVYTRGIGSISNALKQNLSPLEKLVCMQKQ 2256

146 EDTSDSDGLTLK---LQNLKRRKEQEPYERSLRSQ-----NRLKKRR 190

2257 QGNANSTWETWETSCSGVNBALTPSRQDDPTWMTIRFSLRPNAVTTSSQGNRLTKKR 2316

191 --EBDSATGVSQVQSAQDRPLCKQ-----EPAS 219

2317 SKNDVVALGEOKOSQLEHRKELKNTLRKSLRLERNLQSEIHEDVTKVQRHVRPLSN 2376

QY 220 QGVSQSETD-DIENQLEKATQNTENPRE-AGKPKP----- 256
DB 2377 ASPDSQSNERSGEPNLPFKTE---VQNPFRGAPRPKLTKKKEKYCICRTPYDDTKF 2433
QY 257 -----ECG-VYDPNALYCICRQPHN-NRPMTC 282
DB 2434 YVGCDCLSNWFHGDVSTTBASKSLSEFTICDGRARETQOLYSCRPYDSOFTYC 2493
QY 283 DRCEMFHGDVCSISBARGRLLEKNGEDYICPNTTILQVDETNSTATN 331
DB 2494 DCCQDFHRCVGLIQSRAEFL-----DEVCEPC-----QRNDNMAAN 2533
RESULT 3
ID 09MOT1 PRELIMINARY; PRT: 2669 AA.
AC 09MOT1;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-OCT-2002 (TRENBLREL. 22, Last sequence update)
DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)
DS CG32346-PA.
GN E(BX) OR CG32346.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.B., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaniatis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutcliffe G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Chang M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abrial J.F., Agbayani A., An H.J., Andrews-Plambeck C., Baldwin D.,
RA Ballew R.W., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borova D., Borchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mouton G., Milshina N.V., Mobarry C., Morris J., Mostrefi A.,
RA Mount S.M., Moy M., Murphy B., Mobarry C., Morris J., Mostrefi A.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodruff, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Glibbs R.A., Myers E.M., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker S.B., Prochink S.E., Smith C.D.,
RA Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celinker S.E.,

RA Clamp M.B., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.U.,
RA Harris N.L., Kromann B., Marshall B., Millburn G.H., Richter J.,
RA Russo S., Searle S.M.J., Smith E., Shu S., Smutnick P.,
RA Whitefield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,
RA Lewis S.B.,
RT "Annotation of Drosophila melanogaster genome.",
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL, AF003467; AAL1431.1; -
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000637; AT_hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR004022; DPT_dom.
DR InterPro; IPR006209; EGF_1ike.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02178; AT_hook; I.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DPT; 1.
DR Pfam; PF00628; PHD; 3.
DR PRINTS; PRO0503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM0249; PHD; 3.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 2.
DR PROSITE; PS50016; ZF_PHD_2; 2.
SQ SEQUENCE 2669 AA; 300684 MW; 261FF7B7CEBF781B CRC64;
Query Match 7.1%; Score 228; DB 5; Length 2669;
Best local similarity 22.1%; Pred. No. 9.1e-07;
Matches 77; Conservative 52; Mismatches 98; Indels 122; Gaps 14;
QY 93 DSSEPTSTVTDV-ETASGSGVSSSSSEIRSGVSDSL-----GKEHPASSEKAKGGEER 145
DB 2217 EHDEPTNLAGDIDSETDLENKQNESFVVTGVIQKISNALKQGNLSPLEBKJVCWQKQ 2276
QY 146 EDTSDSDSGLTJLKE--LQNRLLRKRGQRPVERSLRSGQ-----NRLRKRR 190
DB 2277 QENANSTNEMETCSGRGVNEBALTPSRQTDTEWKIRTSLLRPNMTTSSQPNRLKQNR 2336
QY 191 --BEDSAFTSGVQIGSAEDRPLCKQ-----EPBAS 219
DB 2337 SKNDEVALGKQKQSLERHKLKKNLLRKSLLERLQSEIHEDVTKVQVRYRPLSN 2396
QY 220 QGPVQSSTED-DIENQLEGKATQGNTEENPRE-AGKPR----- 256
DB 2397 ASPDQSEHNSRSGRPNLDFKYTE--VQNPRIHAGRPFKLLTRKKKLYCICRTPYDDTKF 2453
QY 257 -----ECG-VYDPNALVYICGPHN-NRFMCC 282
DB 2454 YVGCDCLSNMFGDCVSTTEBASKKLSFICIDCKRARBTOOLYCSQNPDESQFYICC 2513
QY 283 DRCEWFHGDGVGISBARGLLERNGEDYICPNCTILLOVODETNGSATN 331
DB 2514 DKCDWPHGRGCVGLQSEAFI-----DYVCEEC-----QKNDANAN 2553
RESULT 4
Q95VB8

ID Q95VB8 PRELIMINARY; PRT; 2669 AA.
AC Q95VB8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Nucleosome remodeling factor large subunit NURF301.
GN E(BX) OR NURF301 OR CG7022 OR CG10894 OR CG17135 OR CG32346 OR CG32478.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21468389; PubMed=11583616;
RA Xiao H., Sandilczopoulos R., Wang H., Hamiche A., Ranallo R., Lee K.,
Fu D., Wu C.,
RT "Dual functions of largest nurf subunit nurf301 in nucleosome sliding
RT and transcription factor interactions.",
RL Mol. Cell 8:531-543(2001).
DR EMBL; AF417921; AAL16644.1; -
DR FlyBase; FBgn000541; R(bx).
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000637; AT_hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR004022; DPT_dom.
DR InterPro; IPR006209; EGF_1ike.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02178; AT_hook; I.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DPT; 1.
DR Pfam; PF00628; PHD; 3.
DR PRINTS; PRO0503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM0249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 2.
DR PROSITE; PS50016; ZF_PHD_2; 2.
SQ SEQUENCE 2669 AA; 300687 MW; 6B4925AFFF489D6F CRC64;
Query Match 7.0%; Score 225; DB 5; Length 2669;
Best local similarity 22.3%; Pred. No. 1.5e-06;
Matches 78; Conservative 53; Mismatches 96; Indels 122; Gaps 15;
QY 93 DSSEPTSTVTDV-ETASGSGVSSSSSEIRSG-----PVSDSL-----GKEHPASSEKAKGGEER 145
DB 2217 EHDEPTNLAGDIDSETDLENKQNESFVVTGVIQKISNALKQGNLSPLEBKJVCWQKQ 2276
QY 146 EDTSDSDSGLTJLKE--LQNRLLRKRGQRPVERSLRSGQ-----NRLRKRR 190
DB 2277 QENANSTNEMETCSGRGVNEBALTPSRQTDTEWKIRTSLLRPNMTTSSQPNRLKQNR 2336
QY 191 --BEDSAFTSGVQIGSAEDRPLCKQ-----EPBAS 219
DB 2337 SKNDEVALGKQKQSLERHKLKKNLLRKSLLERLQSEIHEDVTKVQVRYRPLSN 2396
QY 220 QGPVQSSTED-DIENQLEGKATQGNTEENPRE-AGKPR----- 256
DB 2397 ASPDQSEHNSRSGRPNLDFKYTE--VQNPRIHAGRPFKLLTRKKKLYCICRTPYDDTKF 2453
QY 257 -----ECG-VYDPNALVYICGPHN-NRFMCC 282
DB 2454 YVGCDCLSNMFGDCVSTTEBASKKLSFICIDCKRARBTOOLYCSQNPDESQFYICC 2513

QY 283 DRCSEPHGDCVIGISBARGRLERNGBDYICPNTILQVQDETNGSATN 331
DB 2514 DKCQWPFGRGVIGLQSAEAFI-----DEVCEPC-----QRKNDAMMAN 2553

RESULT 5

Q818F0 PRELIMINARY: PRT: 542 AA.

AC Q818F0 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DS Chromosome scaffold protein p85.

OS Euploceas crassus.
OC Eukaryota; Alveolata; Ciliophora; Spirotrichea; Hypotrichia;
OC Euploidea; Euploidae; Moneuploetes.
NCBI_TaxID=5936;

RA SHARP S.I., Pickrell J.K., Jahn C.L.;
RT "The identification of a novel chromosome scaffold protein that
RT associates with Tec elements undergoing en masse elimination in
RT Euploceas crassus."
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY155457; AAN75020.1; -
SQ SEQUENCE 542 AA; 60264 MW; FOAB3B5E3AA771A1 CRC64;

Query Match 6.5%; Score 210.5; DB 5; Length 542;
Best Local Similarity 18.8%; Pred. No. 2.1e-06;
Matches 111; Conservative 100; Mismatches 195; Indels 183; Gaps 21;

4 KGHLSNEAPRAIKPTSKPRKTWGR-----RTTIKRR 38
3 KKKATATKAGKGLTKKTKTKTAESKTEKATKOVEMKPNPDEEDINKESKTSHTNKS 62
39 GAGTEADPSEEQPOO-----HNLILRRSGROPKTERVEEFLTVRRRGK----- 84
63 EKSDVEHNDADAPQEVKTRKTSRKTKAPKPKTKK-ENNVSGKKTRKYSKKEPDE 121
85 -----KNVPVSG-----LEDSEPTSSVTVDVETASBESVSSSSIRSGPVSISGKHPASS 136
122 EVSEDKYNAKSSSILRRSSRLAANAANKLSAKKENQNSD-----EVEEDTPPKKS 174
137 EK-AKGGSEEDTSDSDGTLTKLQNRKRRKQEPVERSLRSGQNRKRRREEDSA 195
175 KROAKKASKDESSSABBBESSEKSPKSKSKATKPKKVV-----KPTAAKTA 226
196 BTGSVOISAEQDRPLCKQEPASQGPVSQSESTDIDENQLEKATQNTENPREAKPK 255
227 KTKIKR-----EDDE-----SDSSQSGEGEBQBEAASQDQNSG----- 264
256 PECEVYDNPALYCIQPHNNRPMICDRCBEMFHGDCVIGISBARGRLERNGBDYICPN 315
265 -----EGEG-----EENDE----- 273
316 CTILQVQDETNGSATNBODSGRSGVAGDGTCTISIGYKQSGEONGIKRIEKANPSG 375
274 -----EMSEBAGBEBQEDBDMSSBEBGBE-----BEYKQKGGKGVKEBKVAG 322
376 KK-----XKIKFP-----VVEAPG-----APKICIGPCCSSVAPQSPVVCNSDILMH 418
323 KKKOSPKKKKOSPKKKKOSPKKEDVYKKVSTKKNKSKDKGGAKEKKEVAKD--KKD 380
419 AAATRRPLISGKEQKTRPKPKVTK-----PKKSLPKKSGVQVGIYSVHKRLASEKEN 474
381 ATKQKQDEKQKKEITTKQKDKATKDKKDKIOKEKKTSPKSKXKAGTK-----EKKKASQSEKKK 437
QY 475 PVKVMNLASRSETSGEAAACBSSTPSMAHDHYNAVKPEKPKFALSP 523
DB 438 DVKSPKDKDQATSKGKA-----BSKEKQDSPKDKKDKKKISP 477

RESULT 6

Q9W352 PRELIMINARY: PRT: 663 AA.

AC Q9W352 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DS CG17446 protein.

OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;

RA SEQUENCE FROM N.A.

RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinkner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Chape M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abiril J.P., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Borchman M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Glodok C., Gabrielian A.B., Gang N.S., Gelbart W.M., Glasser K.,
RA Fioder A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hosten D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
RA Jaislin M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasbo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mielstina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusekern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Strong R., Sun E.,
RA Spier B., Spradling A.C., Stapleton M., Skupski M.P., Smith T.,
RA Svitakas R., Tector C., Turner R., Venter G., Wang A.H., Wang X.,
RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
DR EMBL: AE003446; AAF6483.1; -
DR P1base: P690030121; CG17446.
DR GO:GO:0003677; F:DNA binding; IEA.
DR GO:GO:0008270; F:zinc ion binding; IEA.
DR GO:GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR002857; znf_CXXC.
DR InterPro: IPR001965; znf_PHD.
DR Pfam: PF00628; PHD; 1.
DR Pfam: PF02008; zf_CXXC; 1.
DR SMART: SM00249; PHD; 1.
DR PROSITE: PS00139; ZF_PHD_1; 1.
DR PROSITE: PS0016; ZF_PHD_2; 1.
SQ SEQUENCE 663 AA; 76696 MW; 8573CD209A828157 CRC64;

Query Match 6.5%; Score 209; DB 5; Length 663;
Best Local Similarity 21.3%; Pred. No. 3.3e-06;
Matches 83; Conservative 39; Mismatches 111; Indels 156; Gaps 15;

146 EDTSDSDSLTKE-----LONLRKREKQEPVERSLRGSQNLKRRKREDSAEFGS 199
 719 -----LYIEVVRVPLTQDLRQGLHAIWQDHTVASQ-----QQQQQPPQDAGGT 764
 200 VOIGSAGDRPLCKQEPF-----ASQG-----PVSGSEF-DDIE 232
 765 TNPDAAGQ-----VQPPQWISGIGITVSGSGQTPPAVGVGYCYFGQOIAASQADDDAH 819
 233 NQLESGKATQG--NREBNREAGKPKPECEVYDPMALYCIQRPNNRMFICCDRCCEMFP 290
 820 SAISSSSMGLASTDIDPEGETEYAPEARAEDSVTRCICELTHDDGTICCDKCSAMOH 879
 291 GDCVGISARGLRLERNEDYICPNCTILQVDETNCSATNEOD-----SQG 337
 880 VDCMGIDR-----QNIPEBTMCELCOPRAV-DKARARALQKQKKEHLVAITQAANGA 932
 338 RSVGADGDTCTSIGTVEKXSGEDQ-----GIKGR1-----EKAANDSGKKK 378
 933 AAVALGTTLSGGLSGGLPMSBELQRLASGLNGCPATGTGMSKKKTKTENSGSTTLKK 992
 379 LKITQPVREAGPARCICPGCSSVAPGPSVYCSNCLKHAATMRFLSSGKEQTKPKPE 438
 993 TK--KSAVGMGEKKNASGSGTPTG-----SSGKTSKSKSR 1026
 439 KVKT 442
 1027 KSKS 1030
 RESULT 8
 08T3Y1 PRELIMINARY: PRT: 366 AA.
 AC 08T3Y1: 09W533:
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DB AT26187P (CG17440 protein).
 DB CG17440.
 OS Drosophila melanogaster (Fruit fly).
 CC Buxarotia, Metazoa; Arthropoda, Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 NC NCB1 TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Friese E.,
 RA George R., Gonzalez M., Guartin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.B., Rubin G.M.,
 RA Ceiniker S.,
 RA Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blake J.R., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.P., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Balles R.M., Baer A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu B., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Doonan K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferrara C., Ferris S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Idegwam C.,
 RA Jaisil K., Kalush F., Karpen G.H., Ke Z., Keanison J.A., Ketchum K.A.,
 RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulj D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mileshina N.V., Mobarry C., Morris J., Mostrelti A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Ceiniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreria S., Friese E., Galle R.F., Gary N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Idegwam C., Jaisil K., Kruse D., Li P., Mattei B., Mostrelti A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,
 RT "Sequencing of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Mistra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Ceiniker S.B.,
 RA Clamm M., Drysdale R., Emmert D., Friese E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Seale S.M.J., Smith B., Shu S., Smutniak F., Whitfield B.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Flybase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY069443; AAL90181.1; -
 DR EMBL: AB003446; AAF46482.2; -
 DR Flybase: FBgn0030120; CG17440.
 DR GO: GO:0003677; P:DNA binding; IEA.
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro: IPR001965; Znf_PHD.
 DR Pfam: PF00628; PHD; 1.
 DR SMART: SM00249; PHD; 1.
 DR PROSITE: PS01359; ZF_PHD_1; 1.
 DR PROSITE: PS50016; ZF_PHD_2; 1.
 SQ SEQUENCE 366 AA; 42048 MW; CB83AAFEAC514CF3 CRC64;

Query Match 6.2%; Score 200; DB 5; Length 366;
 Best Local Similarity 28.3%; Pred. No. 6; 7e-06;
 Matches 54; Conservative 22; Mismatches 65; Indels 50; Gaps 7;

```

QY 267 YCICRQPHNNRPMICCDRCERWPHGDCVIGISEARGLLERNGEDYICPNC-----TI 318
DB 40 YCICRSSDCSFMIGCGCEEMWHDCTEITEKOAHL-----KNYYCRCKKKNPELOTI 95
QY 319 LQVDETNGSATNEOD-----SGCRSVAGDGTCT-----SIGTVEQSGED 360
DB 96 FRUVAATERAASNAASSTLNPAGVPSGAAPAAAPVAATITSSQAAPPTTAARAKKNSA 155
QY 361 QGIGRIEKAANPSGKKLK-----IFQPV-----EAPGAPKICGECSSVA 403
DB 156 QEPK---ESQPTQAGTKKDKAAPTNSVQSPRAVSPFIFLNPLOGIQOCHGPMCCSHA 212
QY 404 QPDVSYCCND 414
DB 213 RPSQRYCSDEC 223

```

RESULT 9

```

ID 074508 PRELIMINARY; PRT: 424 AA.
AC 074508;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE SPCC594.05C.
GN SPCC594.05C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Rieger M., Lyne M., Rajandream M.A., Barrell B.G.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL031523; CAA20664.1; -.
DR PIR; T41449; T41449.
DR GeneDB; Spombe; SPCC594.05C; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS0016; ZF_PHD_2; 1.
SQ SEQUENCE 424 AA; 48683 MW; 2963605C3DFCC0B9 CRC64;

```

Query Match 6.1%; Score 198.5; DB 3; Length 424;

Best Local Similarity 24.1%; Pred. No. 1e-05; Mismatches 139; Indels 115; Gaps 18;

Matches 95; Conservative 46; Mismatches 139; Indels 115; Gaps 18;

```

QY 89 VSLSDSSPTSTVD--VETASGVSSESSSIRSGPVDSLGRKHPASSERAKGGEER 146
DB 20 VKPERSNR---CTIDPHITANNEEKDANI-----LNKSVKQVEEVNG---HV 65
QY 147 DTSDDSDGLTLKELQNRRLRRRQEPYERSLRSGQNRRLKRRREDSAEATGSGAS 206
DB 66 DSSSTETD-IEMQVIQOPTIPKK--PVSARHRGP---KKRGNANS---QLNLSTAD 114
QY 207 QDRPLCKBPSPASQSPVOSSETDDIENQLEGKATQNTENREARNGKPRDEKVVDPN 266
DB 115 HGRP-----L 119
QY 267 YCICRQPHNNRPMICCDRCERWPHGDCVIGISEARGLLERNGEDYICPNCITLQVODITN 326
DB 120 YCICRQPHNNRPMICCDRCERWPHGDCVIGISEARGLLERNGEDYICPNCITLQVODITN 326
QY 327 GSATNEQSGCRSVAGDGTCTSIGTVEQSGED---QGIKRIKKAANPSGKKLKIFQ 383
DB 171 GITTAKR--CTLRECSNPTRENSNYCSDKGVDFREKVK---LSTVEPSATKQLVIFA 225
QY 384 PVEARGAPKICGECSSVAQPDVSYCCNDITLKAATAATMRPLSGKGGKTKPKKVKTK 443

```

```

DB 226 KKRSEFQULGVTVGPRPSQVPEPVY-----NFEIEBANRLANIEIQLNKEKVAASNK 278
QY 444 PEKSLPRCSVQVGKISSVHKRLASERENPVKK 478
DB 279 -----KIFQL-IKSSRRRAVLAKEREG-IKK 304

```

RESULT 10

```

ID Q8VDN7 PRELIMINARY; PRT: 645 AA.
AC Q8VDN7;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein (fragment).
GN PALZ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC021489; AAH21489.1; -.
DR MGI; MGI:244008; Palz.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR006209; BGF-like.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00633; BROMODOMAIN_2; 1.
DR PROSITE; PS01186; BGF_2; 1.
DR PROSITE; PS01186; BGF_2; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS0016; ZF_PHD_2; 1.
DR Hypothetical protein.
FM NON TER
SQ SEQUENCE 645 AA; 72167 MW; 6486F34F339820F9 CRC64;

```

Query Match 6.1%; Score 196.5; DB 11; Length 645;

Best Local Similarity 20.1%; Pred. No. 2.4e-05; Mismatches 125; Indels 113; Gaps 13;

Matches 76; Conservative 64; Mismatches 125; Indels 113; Gaps 13;

```

QY 43 TRADSEQPOQHNLSTR-----SGROPKTERVEFLTYRRRGKKNVPSL-EDS 94
DB 146 TVLSGSCFQPOQVAVAVQLQGVVLSQIQGVVAQIAQSGVAVQIKQLQVPIQVQNS 205
QY 95 SEPTSTVT-----DVETASGVSSESS--SIRSGV----- 124
DB 206 AAQTSVVTVQAASVQBLQRYVQLRDQQQKKQKQIIFERREHTLQASNSGSEIIQGVYVK 265
QY 125 -----SDSLGKHPASSERAKGGE-EDDT 148
DB 266 HNAVIEHLKQKMTTPAREBNQRMVGNQVWKYILDKIKERQDAKKRKEEVEYQKR 325
QY 149 SPSDDGTLT-----KELQNRRLRRR-----QEPYERSLRSGQNRRLKRRRR 192
DB 326 SKQNSAKSLALLFKKKEQLKELIKRLALDLKELQIQQVBSLKRLL-----MKTERENA 380
QY 193 DSARTGVSQIGSABQDRPLCKQBPASQSPVOSSETDDIENQLEGKAT-----QNT 245
DB 381 QAVQANAAASVTPSPAPVAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 440
QY 246 ENPREAGKPRKPC-----EYVDNALYICRQPHN-NRPMICCDRCERWPHGDCVIGIS 298
DB 441 EERKSKSKKKMMISTSKAKQDTRLYICTPYDSKFTYIGCDRCQNMWYHRCVGIHQ 500

```

299 ARGILLERNGEDYICPMC 316
501 SEADLI---DEVCPQC 514
RESULT 11
ID 07YOM3 PRELIMINARY; PRT; 2492 AA.
AC 07YOM3
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DB ATRX protein.
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pongo.
OX NCBI_TaxID=9600;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=22763540; PubMed=1277533;
RA Kltano T., Schwarz C., Nickel B., Paabo S.;
RT "Gene Diversity Patterns at 10 X-Chromosomal Loci in Humans and
Chimpanzees";
RL Mol. Biol. Evol. 20:1281-1289(2003).
DR EMBL; AB102643; BAC01112.1;
SQ SEQUENCE 2492 AA; 282614 MW; 48EC97106D0C9P20 CRC64;
Query Match 5.9%; Score 191.5; DB 6; Length 2492;
Best Local Similarity 18.9%; Pred. No. 0.0028;
Matches 108; Conservative 114; Mismatches 199; Indels 151; Gaps 24;
3 DKHLNBARAKATPTSKPRKTTAKREGADTADSEOPQOHNLSTARS 62
972 DOSDETSHDDKQSKGTBEKKTSDFKKCVI-KMEQYSSSDTEKLPREERIC- 1026
63 GROPKTRVEFLTVRRRGKQNVSLDESEPTSTVTDETASBG---SVESSEH 119
1027 -HPEKGIQIKGTTGDKRKKK-----IRDTSKKKKSLYAKSTGKSCSSBK 1080
120 RSGPVSDSLGKHPASSBRKAKGEBEEDTSDSDGLTKE-----LONRLR- 167
1081 KS-KNGAYGKRRKCTLLGKSSRRQDCSSDTEKYSKEDGCSNDRKLRIELRR 1138
168 ---KREGEPRYKSLRG---QNRLLKRRREDAETGSGVIGAEQDRPLCKGEPAS 219
1139 NLSSGRNTKEIQSGSSSDAEBSSEDNKKKQRTSSKKKAVIV--KCKRNSLATSTYK 1196
220 QGPVQSFTDIDENQLEGATQNTENPRBAKPECEVYDPNALYICRQPHNNRPM 279
1197 QADITSSSSDIEDDDNSIGSSSD--QIKYVTEMLVLSHRTGFC----- 1242
280 ICDRCBEMFHDDCVGISBARGRL-----RNGEDYICPNCTILQVO---DETN 326
1243 ---QSSGD-----EALSKSVPTVDDDDNDPENRIAKWMLERIKANLSDRD 1289
327 GSATWEGDSCRSVAGDGTCTISITVQKSGEDGIGRIKIRKANPSGKKLKIPOPV 386
1290 GSSDDEPEEGKRTKON-----EENPDDEBA-KNQVNSESDSDSESKK- 1333
387 BAPGAPKICIGPCSSVAQPDVYCSNDCTILKHAATMRPLSGKEQKTPKKE--KVKTYP 444
1334 ---PR-----YRHLRLH-KLTVSGSGEGRKTKPKKHKXVKGKN 1370
445 EKFLPKCSVOYGIKISSVHKRLASBKQENPVKVMLASRSR-----TSGEACESSTP 499
1371 RR-----KVSEDSR-DSDPQSGVSESESDQRPRTSKAKKALEENQR 1417
500 SW-----ASDHYNNAVKPEKPEK 517
1418 SYKKKKRRRIKVOBDSSEKNSSEBEEK 1449

RESULT 12
ID 096G15 PRELIMINARY; PRT; 569 AA.
AC 096G15
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DB Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC009449; AA09449.1;
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 569 AA; 62857 MW; 253856PC544PDBEA CRC64;
Query Match 5.8%; Score 188.5; DB 4; Length 569;
Best Local Similarity 21.6%; Pred. No. 7.3e-05;
Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;
28 GFRRT-----IAKREGADTAD-----PSEQOQOHNLSTARSGRPKTERVBEFL 76
52 GKRRTPALKSVSRKARKASDLDQASVSPSEHNSSESEKTSDDPTPEK--KAA 109
77 TVRRRG-----KKNVPSLEDSSEPTSTVTDETASRGVSSESRIRSGPVSDSLK 130
110 VAPAPRGPLGKRRKKKAPASDSDKADSKAPVAVAAASASSSSSSSSSDSVSYK 169
131 EHPASS-----KAKGEEED-----TSDSDGL-TKELONR-----LIR 167
170 KPPRRKPEKPLPPRGKPKPRPPSSSSSDSDSDVDRISEWTRDEARRRLEBAR 229
168 KROEPVRSLSGQNRLLKRRREDSAEFGSVOGSA-----EODRPLCKOPEA- 218
230 RROEELRLRLBOEKEBERRRER--ADRGAEKSGSSGSDDELREDDPEVKKRGRK 287
219 SGPVQSFTDIDENQLEGATQNTENPRBAKPECEVYDPNALYICRQPHNNRPM 276
268 GGPSSSDSR-EAELR-----REAKKSARKPQSSSTP-----ARPD- 325
277 RFLICDRCEWFHDDCVGISBARGRLLENGEDYICPNCTILQVODETNQATNEQDSG 336
326 ---GQKEKVR-----PEKQO 339
337 CSVAGDGTCTISITVQKSGEDGIGRIKIRKANPSGKKLTK-----IFQPVVEAR 390
340 APVAVERT-----RKSEGFSSMDKVKKEKPSVEKQLKHSIKALKVQSDP 390
391 APKCIGPCSSVAQPDVYCSNDCTILK--AAATMRPLSSGKEQKTPKPEKVTYKERS 448
391 VRCL-----NALEHGTILQVTSILQKTDVATLK-----KIRYKANKDVMKAAEY 441
449 LPKGSVOYGIKISSVHK-RLASKEENPVKVMLASRSSTSGKEAACSGSTPSMAD 504
442 TLKSRVLGPKTEAVQVKAKMEKAEKL---AGBELAGEBAPOEFAEDKPSSTD 495
RESULT 13
ID 09BW08 PRELIMINARY; PRT; 670 AA.
AC 09BW08
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DB Similar to hepatoma-derived growth factor, related protein 2.
OS Homo sapiens (Human).


```
QY      85 KNPVSLSDSSEPTSTVTVDVETASBGSVSSSEIRSGP----- 123
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2344 SGLSPGQSQVQTTSQPIPIQPHSLQIPESQGPQSVVMGNNAVIBHLKQKSWTP 2403
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 124 -----VPSLCKEHPASSEKAGGE-REDTSDSDGLTL----- 158
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2404 AERBENQMTVCNQVMKYILDKIDBEKQAKKRRRESEVEQKRSKQATYLSALLFNHK 2463
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 159 KELONRRLRRKE-----QEPVERSLRSONRLKRRREBDSAFTGSAVQIGSAEQD 208
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2464 EQLAAILKRRALADKDLQIEVQZELKRDL-----KTKERDLMQLAQATAV----AAPC 2514
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 209 RPLCKQZPEASQGFVSQSESTDIDEN-----OLEKATQGNTEBNPREAGKPE---- 257
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2515 PPTVPAPAPAPAPPPPPPPAVQHTGLSTPTLPAASQKRRREBKSSSKKKKMS 2574
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 258 ---CEVTDPNALYICICROPHN-ANPMICCDRCCEWTHGDCVGISBARGLLERNGEDYIC 313
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2575 TTSKRTKQDTYLCICTPYDESKFYIGCDRCQWYHGRVCVGLQSEAEI-----DEYVC 2630
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 314 PNC 316
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2631 PQC 2633
```

Search completed: April 20, 2004, 21:19:50
Job time : 48 secs

444 P8K5LPKCSVQWQGIKISSVHRKLASBKRENIPVK 418
| : : | | | : : : : |

neurofilament probb
natural killer cell
hypothetical protee
hypothetical protee
IGa-specific metha
hypothetical protee
DNA-binding protei
hypothetical protee
hypothetical prote
Myd116 protein - m
related to glucan
ES43 protein homol
ES43 like protein
hypothetical protee
natural killer cell
neurofilament probb

QY 359 ED-----OGIKRIEKAAMPSCKKLKIPOVVEARCA-PK 393
 DB 3314 BESRRBSVAERKSPLAYKEASPPASVAESI KDBAEKSKESRRBSVAERKSPLASKEASRPT 3373
 QY 394 CIGCGCCSVAPDVSVCNDCIL-----KHAATWPLSSGKEOKTKPEKVKYTKBEK 446
 DB 3374 SVASVVDDEAKSKESRRDSVAERKSPLASKEASPPASVAESVODAEKSKESRRBSVA 3433
 QY 447 FSLPKCSVQVQIKISSVHKRLA--SEKRENPVKYKMLASRSETSGKACASSTTSMASD 504
 DB 3434 EKSPPLASKEAS-RASVASVYDADAEKSKESRRBSVAERKSPLASKEASPPASVAESVD 3492
 QY 505 HNNYNAVPEKPEKPTALSPILLSKCTYHP 533
 DB 3493 EAEKSKESRRBSVAERKSP-LPSKEASRP 3520

RESULT 12

T29757

protein UNC-89 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999

C/Accession: T29757

R/Du, Z.; Le, T.T.; Wilson, R.

Submitted to the EMBL Data Library, May 1997

A/Description: The sequence of C. elegans cosmid C09D1.

A/Reference number: Z20679

A/Accession: T29757

A/Molecule type: DNA

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Residues: 1-6642 <DUZ>

A/Cross-references: EMBL:AF003131; PDB:1AB54132.1; GSPDB:GN00019; CESP:unc-89

A/Experimental source: Bratton Bristol N2; clone C09D1

C/Genetics:

A:Gene: CESP:unc-89

A:Map position: 1

A/Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6

/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Query Match 5.1%; Score 165; DB 2; Length 6642;

Best local similarity 19.9%; Pred. No. 0.65;

Matches 123; Conservative 87; Mismatches 214; Indels 194; Gaps 24;

QY 8 SNEEAPAIKPTSKERKTGFRRTTAKRGAGDTE--ADPSRQPOQHNLISLRGR 64

DB 1358 ABEELPKEVINDSKRRKKS---DSPDKERKSPKTEKRPASPTKKTGEVKSPEKSPA 1413

QY 65 QPKTER---VEEPLTVRRRKKXNVPLSEDS--SEPTSTVTVE-----TASEGS 112

DB 1414 SPTCKEKSPAAEBVKSPKTKKESPSPTKCKESPSPTKCTGDEVKESPKKSPKTKKES 1473

QY 113 -----VSSSSIRSGPVSLSGKEHPASSEKAKGBEEDT- 148

DB 1474 PEKPEDEVKSPVCKEKSPDATNIVVSSSETTLEKTETTTTEHSEBSRFSVKKEKTPR 1533

QY 149 -----SDSDSDGLTAKELQNRKRRKRGEPVRS-----LRG 180

DB 1534 KVDEKPKSPPTKDKSPKSKITETIKSPVCKEKSPKVEKRPASPTKCKESPEKSPKTK 1593

QY 181 SONLRKRRREDSAGTSGVOIGSAEODRPLCKOPEASOGPVS-----OSRTDI 231

DB 1594 SENEVKSPKTKKESPEKSVVE---ELKSPKESPEKADDPKPTKCKESPKKSATDV 1649

QY 232 ENOLEGATQNTENPREACK-----PKPECEVDNALYICICROPHNNRPMICCDRC 285

DB 1650 KSPKTKKESPEKVEKPTSPKTKESSPTKKTDEVKSPKCKG---KSP----- 1694

QY 286 EEWHGOCVGISBARGLLRNGEDYICPNCTIIQVODETNGSATNBDSGCRSVAAGDT 345

DB 1695 -----QTVBEKASPTKCKESPEKSV----- 1715

QY 346 DCTSIGTVEOKSGBDGIGRIEKAAMPSCKKL--KIFQVVEAPGAPKICGPGCSVA 403

DB 1716 -----VEEVKSPKEKSPDEKABEKPKSPKTKKESPEKSAABEVKSP-TKKEKSPKSAE 1768
 QY 404 QPDVYVCNDCILKHAATWPLSSGKEOKTKPEKVKYTKBEKPLPKCSVQVQIKISSV 463
 DB 1769 KPKSPKTKKESPPVQADAEVR-----SPYKCKESPEKVEKSPAS-----PTKKEKT 1814
 QY 464 HKRLASREKBNPVKKVMLASRSETSGKACASSTTSMASD-----PEKBEKPK- 521
 DB 1815 PEKSAABEKLSPKTK-----EKSPSSPTKCTGDEVKES----- 1857
 QY 522 SPTLSKCTYHPKAGFP 539
 DB 1858 SPT-----PKKSPG 1867

RESULT 13

T21430

hypothetical protein F26H11.3b - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000

C/Accession: T21430

R/Barlow, K.

Submitted to the EMBL Data Library, November 1996

A/Reference number: Z19421

A/Accession: T21430

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-510 <WIL>

A/Cross-references: EMBL:Z81515; PDB:1AB04195.1; GSPDB:GN00020; CESP:F26H11.3b

A/Experimental source: clone F26H11

C/Genetics:

A:Gene: CESP:F26H11.3b

A:Map position: 2

A/Introns: 116/1; 147/3; 288/2; 348/2; 392/1

C/Superfamily: bromodomain homology

F:371-426/Domain: bromodomain homology <BRO>

Query Match 5.1%; Score 164.5; DB 2; Length 510;

Best local similarity 22.6%; Pred. No. 0.041;

Matches 70; Conservative 45; Mismatches 100; Indels 95; Gaps 14;

QY 94 SSEPTSTVTVEETASBGSVSSSEIRSGP-----VSDLSGKEHPASSEKAKG 143

DB 76 SSDKTTPIPKKI-TVQKQPV---PIGKPRRRRCADREISEL---AAKPAAYVK 125

QY 144 EEDTSDSDGLTAKELQNRKRRKRGEPVRSLSGQNRKRRREDSAGTSGVOIG 203

DB 126 EVINPADITLGGDTYDV-----KEQKPYE-SIATVSRRRRT-----SAML 167

QY 204 SABODRPLCKB---PEASOGPVSQSTDDIENOLEGATQGN-----TEENPREAG 252

DB 168 KSBODRDPESQSTAPKSKERTSBPPASHVAFTPGSATPHDNLSTEHCTQKIPDS 227

QY 253 KPKECE-----VYDPMALYICICROPHNN- 276

DB 228 KLYIQCELCANWYHGDGVAGVAGQTLIGLEHWSCEBCLBQGRVVDQPLVYCCQKPYDDT 287

QY 277 RFMPCDRCBEMFPGDCVGISBARGLLRNGEDYICPNCTIIQVODETNGSATNBDS 335

DB 288 KFYVGCSCQGMFHPBCVGTTRAA---EQADYNCPACT-----RABGYEASDVS 338

QY 336 GCRSVAGDT 345

DB 339 GSSRSVQLT 348

RESULT 14

151618

nucleolar phosphoprotein - African clawed frog

C/Species: Xenopus laevis (African clawed frog)

C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999

C/Accession: 151618; S57757

R/Cairns, C.; McStay, B.

CC -1- FUNCTION: Putative transcription factor, weakly pro-apoptotic when overexpressed.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; translocates to the nucleus after pro-apoptotic stimuli.
 CC -1- TISSUE SPECIFICITY: Ubiquitous. Expressed at intermediate levels.
 CC -1- INDUCTION: Upregulated during apoptosis.
 CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.emb-emb.ch/announce/or_send_an_email_to_license@emb-emb.ch).
 CC EMBL: AJ238333; CAB48401.1; -
 CC EMBL: AK042474; BAC31270.1; -
 CC EMBL: AK044919; BAC32141.1; -
 CC MGI: MGI:1344352; D1d01.
 CC GO: GO:0005634; C:nucleus; IDA.
 CC GO: GO:0006915; P:apoptosis; IDA.
 CC InterPro: IPR001965; Znf_PHD.
 CC Pfam: PF00628; PHD; 1.
 CC SMART: SM00249; PHD; 1.
 CC PROSITE: PS01359; ZF_PHD_1; 1.
 CC PROSITE: PS0016; ZF_PHD_2; 1.
 CC Apoptosis; Nuclear protein; Zinc-finger.
 CC DOMAIN 162 170 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 CC ZN_FING 265 319 PHD-TYPE.
 CC CONFLICT 45 45 V -> A (IN REF. 1).
 CC CONFLICT 331 331 D -> N (IN REF. 1).
 CC CONFLICT 353 353 I -> V (IN REF. 1).
 CC SEQUENCE 614 AA; 67409 MW; 15ACCS5A8ABF3539 CRC64;

Query Match 99.7%; Score 3218; DB 1; Length 614;
 Best Local Similarity 99.5%; Pred. No. 2,4e-165;
 Matches 611; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDDKHLSENEAPKAIKPTSKFRKRTWGRRTTIAKRBAGDTADPSEQPOQHNLISLR 60
 DB 1 MDDKHLSENEAPKAIKPTSKFRKRTWGRRTTIAKRBAGDTADPSEQPOQHNLISLR 60
 QY 61 RSGROPKTERVEEFLTYRRRKKONPVSLDSSEPTSTVTVDVETASGVSSSSBR 120
 DB 61 RSGROPKTERVEEFLTYRRRKKONPVSLDSSEPTSTVTVDVETASGVSSSSBR 120
 QY 121 SGFVDSLSKKEPSSERKAGGEBEDTSDSDGLTLKELQNRLLRRKEQEPVERSLFG 180
 DB 121 SGFVDSLSKKEPSSERKAGGEBEDTSDSDGLTLKELQNRLLRRKEQEPVERSLFG 180
 QY 181 SONRLRRKRREDSAYTSGAVOIGASQDRPLCKQEPASQGPVSOSETDINOLEGKAT 240
 DB 181 SONRLRRKRREDSAYTSGAVOIGASQDRPLCKQEPASQGPVSOSETDINOLEGKAT 240
 QY 241 QGNTENPREACKPRPECEVDPNALYCI CROPNNRRFMI CDRCEWFGDCVGISBAR 300
 DB 241 QGNTENPREACKPRPECEVDPNALYCI CROPNNRRFMI CDRCEWFGDCVGISBAR 300
 QY 301 GRLLRNEDYICNPCTTIQVODETNGSTNNODSGCRGVADGDTCTISIGTVEQSGED 360
 DB 301 GRLLRNEDYICNPCTTIQVODETNGSTNNODSGCRGVADGDTCTISIGTVEQSGED 360
 QY 361 GGIKRIEKAAPSGKKLKI FQPVVEAPGAPKICIGPCSSVAQPSVYVCSNDILKHA 420
 DB 361 GGIKRIEKAAPSGKKLKI FQPVVEAPGAPKICIGPCSSVAQPSVYVCSNDILKHA 420
 QY 421 ATWRFLSSGKEQXTKPEKVKTKPKESIPKCSVOVGICSSVHKRLASKEKNPVYKTM 480
 DB 421 ATWRFLSSGKEQXTKPEKVKTKPKESIPKCSVOVGICSSVHKRLASKEKNPVYKTM 480
 QY 481 LASRBTSGKEAKCSSSTPSMASDHNHYAVKRPKEKPPALPTLLSKCTYHPKGFPP 540

DB 481 LASRBTSGKEAKCSSSTPSMASDHNHYAVKRPKEKPPALPTLLSKCTYHPKGFPP 540
 QY 541 SHRLGCGLSLSTRVLYGVILVAVSSSLPARSRVDPAGPQVFLPSLNSLGMFLKSCVG 600
 DB 541 SHRLGCGLSLSTRVLYGVILVAVSSSLPARSRVDPAGPQVFLPSLNSLGMFLKSCVG 600
 QY 601 LMLRALSYSPRP 614
 DB 601 LMLRALSYSPRP 614
 RESULT 2
 ID DATE_HUMAN STANDARD; PRT; 1225 AA.
 AC Q9BIC0; O15043; Q9BIC0; Q9BIC0; Q9H4G6; Q9H4G7; Q9NTU8; Q9NTU8;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Death associated transcription factor 1.
 GN DAF1 OR KIA0333.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RC TISSUE=Placenta;
 RA Isogai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Magatsuma M., Hosofri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takehashi M., Chiba Y., Ishida S., Murakami K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saio K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuda Y.,
 RA Niimura K., Iwayanagi T.;
 RT NEDO human cDNA sequencing project.
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.P., Bates K.N., Beard L.M., Beare D.W.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgman A.M., Brown A.J.,
 RA Buck D., Burrill M.D., Butler A.P., Carder C., Carter N.P.,
 RA Chappas J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Graham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.B.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle B., Hunt A.R., Hunt S.B., Jekosch K., Johnson C.W., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA LeHayesialho M.H., Leverisa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McLeay K., McMurray A.A.,
 RA Milne S.A., Mistry D., Moore M.J.P., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.B., Sehra H.K., Showkeen R., Sims S.,
 RA Skues C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 RA Whitehead L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20."
 RL Nature 414:865-871 (2001).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
 RC TISSUE=Colon, Kidney, and Muscle;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

Query	Database	Score	Length	DB 1	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10	DB 11	DB 12	DB 13	DB 14	DB 15	DB 16	DB 17	DB 18	DB 19	DB 20	DB 21	DB 22	DB 23	DB 24	DB 25	DB 26	DB 27	DB 28	DB 29	DB 30	DB 31	DB 32	DB 33	DB 34	DB 35	DB 36	DB 37	DB 38	DB 39	DB 40	DB 41	DB 42	DB 43	DB 44	DB 45	DB 46	DB 47	DB 48	DB 49	DB 50	DB 51	DB 52	DB 53	DB 54	DB 55	DB 56	DB 57	DB 58	DB 59	DB 60	DB 61	DB 62	DB 63	DB 64	DB 65	DB 66	DB 67	DB 68	DB 69	DB 70	DB 71	DB 72	DB 73	DB 74	DB 75	DB 76	DB 77	DB 78	DB 79	DB 80	DB 81	DB 82	DB 83	DB 84	DB 85	DB 86	DB 87	DB 88	DB 89	DB 90	DB 91	DB 92	DB 93	DB 94	DB 95	DB 96	DB 97	DB 98	DB 99	DB 100	DB 101	DB 102	DB 103	DB 104	DB 105	DB 106	DB 107	DB 108	DB 109	DB 110	DB 111	DB 112	DB 113	DB 114	DB 115	DB 116	DB 117	DB 118	DB 119	DB 120	DB 121	DB 122	DB 123	DB 124	DB 125	DB 126	DB 127	DB 128	DB 129	DB 130	DB 131	DB 132	DB 133	DB 134	DB 135	DB 136	DB 137	DB 138	DB 139	DB 140	DB 141	DB 142	DB 143	DB 144	DB 145	DB 146	DB 147	DB 148	DB 149	DB 150	DB 151	DB 152	DB 153	DB 154	DB 155	DB 156	DB 157	DB 158	DB 159	DB 160	DB 161	DB 162	DB 163	DB 164	DB 165	DB 166	DB 167	DB 168	DB 169	DB 170	DB 171	DB 172	DB 173	DB 174	DB 175	DB 176	DB 177	DB 178	DB 179	DB 180	DB 181	DB 182	DB 183	DB 184	DB 185	DB 186	DB 187	DB 188	DB 189	DB 190	DB 191	DB 192	DB 193	DB 194	DB 195	DB 196	DB 197	DB 198	DB 199	DB 200	DB 201	DB 202	DB 203	DB 204	DB 205	DB 206	DB 207	DB 208	DB 209	DB 210	DB 211	DB 212	DB 213	DB 214	DB 215	DB 216	DB 217	DB 218	DB 219	DB 220	DB 221	DB 222	DB 223	DB 224	DB 225	DB 226	DB 227	DB 228	DB 229	DB 230	DB 231	DB 232	DB 233	DB 234	DB 235	DB 236	DB 237	DB 238	DB 239	DB 240	DB 241	DB 242	DB 243	DB 244	DB 245	DB 246	DB 247	DB 248	DB 249	DB 250	DB 251	DB 252	DB 253	DB 254	DB 255	DB 256	DB 257	DB 258	DB 259	DB 260	DB 261	DB 262	DB 263	DB 264	DB 265	DB 266	DB 267	DB 268	DB 269	DB 270	DB 271	DB 272	DB 273	DB 274	DB 275	DB 276	DB 277	DB 278	DB 279	DB 280	DB 281	DB 282	DB 283	DB 284	DB 285	DB 286	DB 287	DB 288	DB 289	DB 290	DB 291	DB 292	DB 293	DB 294	DB 295	DB 296	DB 297	DB 298	DB 299	DB 300	DB 301	DB 302	DB 303	DB 304	DB 305	DB 306	DB 307	DB 308	DB 309	DB 310	DB 311	DB 312	DB 313	DB 314	DB 315	DB 316	DB 317	DB 318	DB 319	DB 320	DB 321	DB 322	DB 323	DB 324	DB 325	DB 326	DB 327	DB 328	DB 329	DB 330	DB 331	DB 332	DB 333	DB 334	DB 335	DB 336	DB 337	DB 338	DB 339	DB 340	DB 341	DB 342	DB 343	DB 344	DB 345	DB 346	DB 347	DB 348	DB 349	DB 350	DB 351	DB 352	DB 353	DB 354	DB 355	DB 356	DB 357	DB 358	DB 359	DB 360	DB 361	DB 362	DB 363	DB 364	DB 365	DB 366	DB 367	DB 368	DB 369	DB 370	DB 371	DB 372	DB 373	DB 374	DB 375	DB 376	DB 377	DB 378	DB 379	DB 380	DB 381	DB 382	DB 383	DB 384	DB 385	DB 386	DB 387	DB 388	DB 389	DB 390	DB 391	DB 392	DB 393	DB 394	DB 395	DB 396	DB 397	DB 398	DB 399	DB 400	DB 401	DB 402	DB 403	DB 404	DB 405	DB 406	DB 407	DB 408	DB 409	DB 410	DB 411	DB 412	DB 413	DB 414	DB 415	DB 416	DB 417</
-------	----------	-------	--------	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	----------


```

QY 220 QGVSQSEETDIENOLEGKATQNTENPREAKRPECEVYDPAALYICRQPHNNFM 279
DB 1197 QADITSSSDIFDDQNSIGESSDE---QKIKVENTVLSSHTGFC----- 1242
QY 280 ICCRCCEWFGDCVGISBAGRL-----BRNEDVICRCTLLOYO---DFTN 326
DB 1243 -----QSSGD-----BALSKSVPTVDDDDNDENNAKQMLLEBKANISSBD 1289
QY 327 GSATNEODSGRVSADGDTCTSIGTVEQKSGEDGIGRIEKAANPSGKKKLIKQPVV 386
DB 1290 GSSDDEPEEGKGRKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 1333
QY 387 EAPAPKICIGGCCSSVAPDSVYCSNDCTLKAAATMFLSSGKQKTKPKK---KVKTKP 444
DB 1334 -----PR-----YRHLRLRH-KLTVSDGSGSEKTKPKKPKKQKQKQKRN 1370
QY 445 EKFLSPKSVGVGKISVHKRLASKEKRNPKYKMLASRS-----TSQKPAACESSTP 499
DB 1371 RR-----KVSSBDE-DSDPQSGVSESVSESEDEQRPRTSAKKALEENOR 1417
QY 500 SM-----ASDHNYNAVPEKPEK 517
DB 1418 SYKQKKRRRIKQVQDSSSENKSNSEBEEK 1449

RESULT 6
ID ATRX_CAEEL STANDARD; PRT; 1359 AA.
AC Q9UT80; 002061;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
GN XNP-1 OR B0041.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodietinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RX MEDLINE=99365296; PubMed=10433961;
RA Villalva L., Fontes M., Ewbank J.J.;
RT "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RT the human XNP/ATR-X gene.";
RL Gene 236:13-19(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fulton R., Wohlmann P.;
RA Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Could be a global transcriptional regulator. Modifies
CC gene expression by affecting chromatin (Potential).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: AF134186; AAD55361.1;
CC EMBL: AF000196; AAC24256.1;
CC PIR: T34036; T34036.
CC WormPep; B0041.7; CE17314.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR000330; SNF2_N.
CC Pfam; PF00271; helicase_C_1.
CC Pfam; PF00176; SNF2_N_1.
CC SMART; SM00487; DEXDC; 1.

```

```

DR SMART; SM00480; HELIC; 1.
DR PROSITE; PS00690; DEAD_ATP_HELICASE; FALSE_NEG.
KW DNA repair; Hydrolyase; Helicase; Nuclear protein; ATP-binding;
KM DNA-binding.
FT NP BIND 496 503
FT SITE 636 639
FT DOMAIN 67 70
FT DOMAIN 266 272
FT DOMAIN 276 281
FT DOMAIN 372 375
FT DOMAIN 603 608
FT DOMAIN 859 862
FT COMPLET 479 479
SQ SEQUENCE 1359 AA; 156191 MW; BB4342547D4F64 CRC64;

Query Match 5.7%; Score 185; DB 1; Length 1359;
Best Local Similarity 21.4%; Pred. No. 0.0098;
Matches 146; Conservative 88; Mismatches 217; Indels 232; Gaps 33;

QY 8 SNEAPKAIKPTSKREPKRTTWGRTTIKRBAGDTEADPSE--QQPQHNLRLRSGRQ 65
DB 210 SESEDEKVKKKKKKKK-----VVKSESSEDEAPKKKTKRKTKTSSEBSSES 261
QY 66 PRTTREVREPLTVRRGKKAVVSLSDSEPTSTVDVETASE---GSV-----ESSS 117
DB 262 EKSDDEEEKSSPPKPKK--PLAVKGLSDSESEBBDVRLPQKKRGAVTLISDSED 319
QY 118 EIRSGVSDSLQKEHPASSERAKGGESEBDTSDSDGLITLKLONLRKRREO-EPVER 176
DB 320 EMDQKSEBSEADVEKVKAKKQSSSESGDSSEGSITV---NRKSKKEKKEKKKK 375
QY 177 SLRSGQNLKRRKREDEAFTGVSIGSABORPLCKQEPKASQGPVSQSETDILENOL 236
DB 376 GIMDSSTKLQETITDAERAE-----KERRKRLKQKQKFN--GIYIASEDLTLEMT 425
QY 237 GKATQNTRE---NP-----REAGKPRPECEVYDPAALYICR--QPHNNFMICC--- 282
DB 426 GTSSRKRLKSVLDDSDSTVDESKKP---VEVH--NLVRLTKHQHAGIQFMWDCAE 480
QY 283 --DRCEWFGH--DCVGISBAGRL-----BRNED-----YICPCTIL 319
DB 481 SLDRLDTEGSGGILAHCMGL-----GKTLQVITFLHTVLMEHKIGCKKRVLVVPPQVIT 536
QY 320 -----QVQPEMGSAITNED-----SGCRSVAGDGTCTSIG 351
DB 537 NMFKEFKQVLDNDEELTITVNEIDSYKTIEDRRALKAMWSKTPSMITGIDLFRIL 596
QY 352 TYEQ--KSGEDGIGRIEKAANPSGKKKLIKQPVVEAPGAPKICIGPCSSVAPDSVY 409
DB 597 TVEDDPKKKKKPKNRRRLKAKED-----FRKYLQNP-----PDVVV 634
QY 410 CSNDCTLKAAATMFLSSGKQKTKPKKQKTKPEKPSLPKCSQVQGIKISSYHKRLAS 469
DB 635 CD-----BAHKLKNDLSLKKQVYI----- 655
QY 470 EKRBNPKYKMLASSTSGKEAACESTPMSADSHNVAVAPKPEKPTALPTLLSKC 529
DB 656 -----LTKRRLITLTP-----LQNNIME--- 674
QY 530 TYHPKAGFGPSEHILGCGISRTVLGVLYIVASSSLPARRYQDASGPV-FLPSILW 588
DB 675 -YHCANVTKP-----GLLG-TKTEPARFVNI-----NGRTVDAFLVSEFKRRRC 721
QY 589 SLSGWFLASCVG-----LMLENI 606
DB 722 HVLVDHLKKCVDRKQYRVLTETAI 744

RESULT 7
ID ATRX_HUMAN STANDARD; PRT; 2492 AA.
AC P46100; P51068; Q15886; Q9H021; Q9NTS3;
DT 01-NOV-1995 (Rel. 32, Created)

```

DT 28-FEB-2003 (Ref. 41, Last sequence update)
 DT 15-MAR-2004 (Ref. 43, Last annotation update)
 DT Transcriptional regulator ATRX (X-linked helicase II) (X-linked
 DE nuclear protein) (XNP) (Znf-HX).
 GN ATRX OR RAD54L OR HX2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOPRIMS 1; 2; 3; 4 AND 5), VARIANTS SER-596;
 RP GLU-740 AND SER-1860, AND VARIANTS ATR-X.
 RA MEDLINE=97123494; PubMed=8968741;
 RA Picketts D.J.; Higgs D.R.; Bachoo S.; Blake D.J.; Quarrell O.W.J.;
 RA Gibbons R.J.;
 RA "ATRX encodes a novel member of the SNF2 family of proteins: mutations
 RT point to a common mechanism underlying the ATR-X syndrome.";
 RT Hum. Mol. Genet. 5:1899-1907(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOPRIMS 2 AND 4).
 RA MEDLINE=97386582; PubMed=9244431;
 RA Villard L.; Losel A.-M.; Cardoso C.; Proud V.; Chiaroni F.;
 RA Colleaux L.; Schwartz C.; Fontes M.;
 RA "determination of the genomic structure of the XNP/ATRX gene encoding
 RT a potential zinc finger helicase.";
 RT Genomics 43:149-155(1997).
 RN [3]
 RP SEQUENCE FROM N.A. AND VARIANTS SER-596 AND GLU-740.
 RA MEDLINE=22763540; PubMed=1277533;
 RA Kitzano T.; Schwarz C.; Nickel B.; Paabo S.;
 RA "Gene diversity patterns at 10 X-chromosomal loci in humans and
 RT chimpanzees.";
 RT Mol. Biol. Evol. 20:1281-1289(2003).
 RN [4]
 RP SEQUENCE OF 860-2492 FROM N.A.
 RA MEDLINE=95179111; PubMed=7874112;
 RA Stayton C.L.; Dabovic B.; Gulsian M.; Gecz J.; Broccoli V.;
 RA Giovannazzi S.; Bosolascio M.; Monaco L.; Raetan S.; Boncinelli B.;
 RA Bianchi M.B.; Consalez G.G.;
 RA "Cloning and characterization of a new human Xq13 gene, encoding a
 RT putative helicase.";
 RT Hum. Mol. Genet. 3:1957-1964(1994).
 RN [5]
 RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
 RA MEDLINE=94214473; PubMed=8162050;
 RA Gecz J.; Pollard H.; Consalez G.; Villard L.; Stayton C.L.;
 RA Milasseau P.; Kiretschatsky M.; Fontes M.;
 RA "Cloning and expression of the murine homologue of a putative human
 RT X-linked nuclear protein gene closely linked to PGK1 in Xq13.3.";
 RT Hum. Mol. Genet. 3:39-44(1994).
 RN [6]
 RP SEQUENCE OF 2401-2492 FROM N.A. AND VARIANTS ATR-X.
 RA MEDLINE=95211835; PubMed=7697714;
 RA Gibbons R.J.; Picketts D.J.; Villard L.; Higgs D.R.;
 RA "Mutations in a putative global transcriptional regulator cause X-
 RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
 RT Cell 80:837-845(1995).
 RN [7]
 RP SEQUENCE OF 1375-2492 FROM N.A.
 RA Pearce A.; Chapman J.;
 RA Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP BZHZ BINDING.
 RA MEDLINE=98167853; PubMed=9499421;
 RA Cardoso C.; Timelst S.; Villard L.; Kiretschatsky M.; Fontes M.;
 RA Colleaux L.;
 RA "Specific interaction between the XNP/ATR-X gene product and the SER
 RT domain of the human BZHZ protein.";
 RT Hum. Mol. Genet. 7:679-684(1998).
 RN [9]
 RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
 RT HETEROCHROMATIN.
 RA MEDLINE=20040663; PubMed=10570185;

RA McDowell T.L.; Gibbons R.J.; Sutherland H.; O'Rourke D.M.;
 RA Bickmore W.A.; Pombo A.; Hurley H.; Gatter K.; Picketts D.J.;
 RA Buckle V.J.; Chapman L.; Rhodes D.; Higgs D.R.;
 RA "Localization of a putative transcriptional regulator (ATRX) at
 RT pericentromeric heterochromatin and the short arms of acrocentric
 RT chromosomes.";
 RT Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
 RN [10]
 RP DISEASE.
 RA MEDLINE=20213147; PubMed=10751095;
 RA Villard L.; Fontes M.; Ades L.C.; Gecz J.;
 RA "Identification of a mutation in the XNP/ATR-X gene in a family
 RT reported as Smith-Petersen-Myers syndrome.";
 RT Am. J. Med. Genet. 91:83-85(2000).
 RN [11]
 RP DISEASE.
 RA MEDLINE=22804448; PubMed=12858175;
 RA Gibbons R.J.; Pellagatti A.; Garrick D.; Wood W.G.; Malik N.;
 RA Ayub H.; Langford C.; Boulwood J.; Wainscoat J.S.; Higgs D.R.;
 RA "Identification of acquired somatic mutations in the gene encoding
 RT chromatin-remodeling factor ATRX in the alpha-thalassemia
 RT myelodysplasia syndrome (ATMDS).";
 RT Nat. Genet. 34:446-449(2003).
 RN [12]
 RP VARIANT ATR-X SER-1713.
 RA MEDLINE=97196774; PubMed=9043863;
 RA Villard L.; Lacombe D.; Fontes M.;
 RA "A point mutation in the XNP gene, associated with an ATR-X phenotype
 RT without alpha-thalassemia.";
 RT Eur. J. Hum. Genet. 4:316-320(1996).
 RN [13]
 RP VARIANT JM GLN-2131.
 RA MEDLINE=96224392; PubMed=8630485;
 RA Villard L.; Gecz J.; Mactel J.-F.; Fontes M.; Saugier-Veber P.;
 RA Munich A.; Lyonnet S.;
 RA "XNP mutation in a large family with Jubb-Berg-Mazidi syndrome.";
 RT Nat. Genet. 12:359-360(1996).
 RN [14]
 RP VARIANTS ATR-X.
 RA MEDLINE=97467722; PubMed=9326931;
 RA Gibbons R.J.; Bachoo S.; Picketts D.J.; Aftimos S.; Aisenbauer B.;
 RA Bergoffen J.; Berry S.A.; Dahl N.; Fryer A.; Keppler R.; Kurosawa K.;
 RA Levin M.L.; Masuno M.; Neri G.; Pierpont M.E.; Slaney S.F.;
 RA Higgs D.R.;
 RA "Mutations in transcriptional regulator ATRX establish the functional
 RT significance of a PHD-like domain.";
 RT Nat. Genet. 17:146-148(1997).
 RN [15]
 RP VARIANT ATR-X LEU-246.
 RA MEDLINE=20123062; PubMed=10660327;
 RA Fichera M.; Romano C.; Castiglia L.; Palla P.; Ruberto C.; Amata S.;
 RA Greco D.; Cardoso C.; Fontes M.; Ragnu A.;
 RA "New mutations in XNP/ATR-X gene: a further contribution to
 RT genotype/phenotype relationship in ATR/X syndrome.";
 RT Hum. Mutat. 12:214-214(1998).
 RN [16]
 RP VARIANT SHS LYS-1742.
 RA MEDLINE=99347960; PubMed=10417298;
 RA Losel A.-M.; Millan J.M.; Villard L.; Orellana C.; Cardoso C.;
 RA Prieto F.; Fontes M.; Martinez F.;
 RA "Mutation of the XNP/ATR-X gene in a family with severe mental
 RT retardation, spastic paraplegia and skewed pattern of X inactivation:
 RT demonstration that the mutation is involved in the inactivation
 RT bias.";
 RT Am. J. Hum. Genet. 65:558-562(1999).
 RN [17]
 RP VARIANT CMS THR-2050.
 RA MEDLINE=99326061; PubMed=10398237;
 RA Abidi F.; Schwartz C.B.; Carpenter N.J.; Villard L.; Fontes M.;
 RA Curtis M.;
 RA "Carpenter-Mazidi syndrome results from a mutation in XNP.";
 RT Am. J. Med. Genet. 85:249-251(1999).
 RN [18]


```

OS Canis familiaris (Dog) .
OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis .
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Heart, and Skeletal muscle;
RX MEDLINE=99428545; PubMed=10497235;
RA Kobayashi Y.M., Jones L.R.;
RT "Identification of triadin 1 as the predominant triadin isoform
   RT expressed in mammalian myocardium."
RL J. Biol. Chem. 274:28660-28668(1999).
CC -1 FUNCTION: May be involved in anchoring caldesmon to the
   CC junctional sarcoplasmic reticulum and allowing its functional
   CC coupling with the ryanodine receptor (By similarity).
CC -1 SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
   CC reticulum.
CC -1 ALTERNATIVE PRODUCTS:
   CC Event=Alternative splicing; Named isoforms=3;
   CC Name=Skeletal;
   CC IsoId=P82179-1; Sequence=Displayed;
   CC Name=Cardiac 1;
   CC IsoId=P82179-2; Sequence=VSP_004001, VSP_004002;
   CC Name=Cardiac 3;
   CC IsoId=P82179-3; Sequence=VSP_004003, VSP_004004;
   CC -1 TISSUE SPECIFICITY: Skeletal and cardiac muscle.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF165916; AAF00222.1; -.
DR EMBL; AF165915; AAF00221.1; -.
DR EMBL; AF165917; AAF00223.1; -.
RW Transmembrane; Sarcoplasmic reticulum; Glycoprotein;
KW Alternative splicing.
FT INIT MET      0
FT DOMAIN        1  46
FT TRANSMEM     47  67
FT DOMAINS       68  700
FT CARBOHYD      74  74
FT CARBOHYD     616  616
FT VARSPPLIC    257  277
FT FT
FT FT
FT VARSPPLIC     278  700
FT FT
FT VARSPPLIC     466  466
FT FT
FT FT
FT VARSPPLIC     524  579
FT FT
FT FT
FT FT
SQ SEQUENCE 700 AA; 78152 MW; F033E3AABEE0C56 CRC64;
Query Match          5.2%; Score 169; DB 1; Length 700;
Best Local Similarity 19.9%; Pred. No. 0.034;
Matches 114; Conservative 84; Mismatches 190; Indels 184; Gaps 24;
OY 3 DKGHLS-----NEBAPKAIKPFSKEFRKTVGRRFTIAKBGAGDTADPSEQF 52
DB 129 DKGIEEPPLAKQXIHKKRKKERKPKTLAK-----VAHREKKVKRKESKSK 179
OY QOHNIILSRSGROPKRTERVEEPIITVRRRGKNVPVLSDESPITSJTVDV----- 105
DB 180 AIHKETIKKKRBPETTKMAEKBRKAKTBEKI KGVKGGQGEKVPTAAKYVBQKTPKA 239

```



```

OY 106 -----ETASGVSSESS-----SEIRSG-----PVSISLCKEHPASSEK 138
DB 240 KEKEGKETAAVAKHQKQOYAFRCRYMIDMFVHGDIARPQSPALPPPLPTVQASRPTPASP 299
OY 139 AKGEBEEDTSDSDGLTLKELQNRRLRKREQEPVERSLSGQNRLRKREBEDSAFTG 198
DB 300 TLBGRBEBEKKAAE-----KVTSE-YYKKEKEDV-----KKKSDKOTAI-- 338
OY 199 SVQISASQDRPLCKQBPASQGP---VSQSETDDIENQLEKATQANTENPREAKPK 255
DB 339 -----DVEKGP--GKAPETKQGTIKVVAQAAKQDEKEDSKTTPVEBHPK--GKKQ 389
OY 256 PECEYTPRALYICIGPHNNRPMICCRCEMFMHDCVGISEARGRLLEGEQYICN 315
DB 390 EKKEKY-----VEPAKSSKKEHSAPESEKQY 415
OY 316 CTIIQVDETNGSATNEQSGCRSVAGDGTCTSIGTEVQK--SGEDQGIKGRIRKAAFPS 374
DB 416 AKTERAKETSAASATKAVPGKK-----BEKTTTVQGEIRKESGTSSTASQKBPB 468
OY 375 GKXKLKIPQVVBAPGAPKICIGPCSSVAQPOSVYCSNDCLKHAATNRPILSSGKEQXT 434
DB 469 IKKDEKM-----PKADKVRP-----KPPQS-----QVKKEEKS 497
OY 435 KPEKVKTKPEKPSLPKCSVQVQIKISVHKLASKEKREPNVKKWMLASRSETSGKBAAC 494
DB 498 ESQVKEKAPQEQ-----DIKKEKTVSHQKPEKVKQVKAT-----EKAAI 539
OY 495 ESST---PSMADHNYN---AVPEKPEKPTA 520
DB 540 EKVTKPKPAKKAHQEKESPTIKTDKP-KPTS 570

```

RESULT 10
LAF4_HUMAN STANDARD; PRT; 1227 AA.

```

AC 01-OCT-1996 (Rel. 34, Created)
DB 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE LAF-4 protein (Lymphoid nuclear protein related to AP4).
GN LAF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96141096; PubMed=8555498;
RA Ma C., Staudt U.M.;
RT "LAF-4 encodes a lymphoid nuclear protein with transactivation
RT potential that is homologous to AP-4, the gene fused to MLL in
RT t(4;11) leukemias."
RL Blood 87:734-745(1996).
CC - FUNCTION: Putative transcription activator that may function in
CC lymphoid development and oncogenesis. Binds, in vitro, to
CC double-stranded DNA.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - TISSUE SPECIFICITY: Preferentially expressed in lymphoid tissues,
CC highest levels being found in the thymus.
CC - SIMILARITY: Belongs to the AP4 family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U34360; AAA98763.1; --
CC TRANSFAC: T04866; --
CC GeneW: HGNC:6473; LAF4.

```

```

DR MIM: 601464; --
DR GO: GO:0005634; C:nucleus; TAS.
DR GO: GO:0007275; P:development; TAS.
DR InterPro: IPR007797; AF-4.
DR Pfam: PF05110; AF-4; 1.
KW Nuclear protein; Transcription regulation; Activator; DNA-binding.
FT DOMAIN 413 419 POLY-SER.
FT DOMAIN 422 432 POLY-SER.
FT DOMAIN 440 445 POLY-SER.
FT DOMAIN 670 679 POLY-SER.
SQ SEQUENCE 1227 AA; 133734 MW; 634B896FD7B9BBE7 CRC64;

```

Query Match 5.2%; Score 169; DB 1; Length 1227;
Best local similarity 19.1%; Pred. No. 0.063;
Matches 144; Conservative 95; Mismatches 209; Indels 306; Gaps 34;

```

OY 13 PRAIKETSEKFRKTGFRRTTAKREAGADTEADPSDQPOQHNLSKRS---GRPKRT 69
DB 133 PAAV--PVQSKKGTGQWQK-----AGHPSPDQGORATQGSILFTLLGQVGRQOPRA 183
OY 70 ER---VEERLTIVRR-----RQKNVPVSL-----SEIRSG----- 91
DB 184 KQVCNVEVGLQTERPPAAMAKHSSGHCVQNPPLSLAKPSLVQKPTATVVRPMDQDQ 243
OY 92 --EDSGEPTSTVTDVETASGVSSESS-----SEIRSG----- 122
DB 244 APDESFKLSSSTSVHCTSYRGVPAKBPAPAKAKLSKFSIPKQGBESGSETNSCVE 303
OY 123 -----PVS--DSLCKEHP-----ASSEKAKGEBEEDTSDSDGL 156
DB 304 EIRREWTMLPPLSAIQAPKEVETKPPFNKQSOLVSSGHNNDKKDABESBDNGTSNT 363
OY 157 TLKELQNRRLRKREQEPVERSLRGSQNRLKREKREBDSARTSVOI--GSAQDRPLCKQ 214
DB 364 SM--LEDDIKLSDEE-----ENQQAQRTALRLSDSAVVOQPPKCR 405
OY 215 EPEASQAPVSQSETDDIENQLEKATQANTER-----NPREAKP-----KPECEYTPDNA 265
DB 406 SVPSGKSSSSSSSGTSSSSSDSESSGSDSEFTSSSSSESGKPPHFSPEAB----- 459
OY 266 LYCICRQPHNNRPMICCRCEMFMHDCVGISBARGRLERNGEDYICPNCITLQVODET 325
DB 460 -----PASSNKQQL-----DKWL-----NKVNPKHPPLTIQNES 488
OY 326 NSGATNEQSGCRSVAGDGTCTSIGTEVQKSGEDQGIKGRIRKAAFPSKKKLTQPV 385
DB 489 HGSBSNOY---YNPVKEDVQDCKVDPVQCPSLREKIKSTCKEBOQRPRTANK----- 538
OY 386 VEAPGA-----PKCIGKCSVAQPDVSVCSNDCLKHAATNRPILSSGKE-----Q 432
DB 539 --APSGKGVKQSPPAVAVAVAAAPPAVPCAP---AENADAPAR--RAGKKPTTRTE 592
OY 433 KTK-----PKKVKTKTP-----EKES- 448
DB 593 RTSGAGDANCHREBPAAADALGTSVVPEPEPKTRPCGNRRASHRKELRSSVTCERRRT 652
OY 449 -----LPR-----CSVQVQIKISVHKRLASKEKREPNVKKWMLASRSETSG-----KBA 493
DB 653 RGLSRIVPSKEPRTETSSSSSSSDLESBEQETPLSAQVVAASASGNDQLKEAA 712
OY 494 CESTSPMSADHNYN-----VPEKPEKPTALSP----- 523
DB 713 ANGSGPRAPVGSINRRATTSDIKALEBOGYTLVPCFGRNELSLPLKSDSIRSLMWKIDL 772
OY 524 TLISKCTTH--PKAGPFC-----PSH 542
DB 773 TLLSRIPEDHPOR--DGVLSAPATKQSBAPSH 804

```

RESULT 11
CGBP_MOUSE STANDARD; PRT; 660 AA.
ID CGBP_MOUSE
AC Q9CWM7;

DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DB Cpg binding protein (Protein containing PHD finger and CXXC domain 1).
 GN CGBP OR PCCL1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic stem cells;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi I., Fukuda S.,
 RA Aizawa K., Izawa M., Mishi K., Kiyosawa H., Kondo S., Yamana I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kanakawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Plechmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Koehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirml L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Oikido T., Furuno M., Aono H., Balderelli R., Barsh G.,
 RA Blake J., Botfield D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guncionich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberters P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszewski B., Yoshida K., Hasegawa Y., Kawaji H., Kohlschütter S.,
 RA Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection.";
 RT Nature 409:685-690(2001).
 RL -1- PUNCTUON: Transcriptional activator that exhibits a unique DNA
 binding specificity for [AC]CpG[AC] unmethylated CpG motifs (By
 similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -1- SIMILARITY: Contains 1 CXXC-type zinc finger.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.1sb-std.ch/announce/>
 or send an email to license@1sb-std.ch).
 CC -----
 DR EMBL: AK010337; BAB26862.1; -
 DR MGI: MGI:1921572; Cgdp.
 DR GO: GO:0016363; C:nuclear matrix; IDA.
 DR InterPro: IPR002857; Znf_CXXC.
 DR InterPro: IPR001965; Znf_PHD.
 DR Pfam: PF00628; PHD; 1.
 DR Pfam: PF02008; ZF-CXXC; 1.
 DR SMART: SMO0249; PHD; 1.
 DR PROSITE: PS01359; ZF_PHD_1; 1.
 DR PROSITE: PS0016; ZF_PHD_2; 1.
 DR Transcription regulation; Activator; DNA-binding; Zinc-finger; Zinc;
 KW Metal-binding; Coiled coil; Nuclear protein.
 KM ZN FING 28
 FT ZN FING 165 212 CXXC-TYPE.
 FT ZN FING 260 321 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 325 363 ARG/LYS-RICH (BASIC).
 FT DOMAIN 426 479 COILED COIL (POTENTIAL).
 FT DNA BIND 156 161 POLY-GLN.
 FT FT 225 BY SIMILARITY.
 SQ SEQUENCE 660 AA; 76166 MW; 9D92CB6DA82069EA CRC64;
 Query Match 5.2%; Score 167.5; DB 1; Length 660;
 Best local Similarity 18.4%; Pred. No. 0.036; 72; Indels 237; Gaps 13;
 Matches 76; Conservative 29; Mismatches 237; Gaps 13;

QY 235 LRGATGANTENREBRAGKRPCECEVVDNALYCIQROPHNNRPMICDRGEWPHGDCV 294
 DB 1 MEG---DGSIDLEPPDAGDSYSENGENAP--IYLCIRPDIINCFMIGDNKMBPHGDCI 55
 QY 295 GISERAGLIERNGEDYICPNC---TLIQVQ-----DETNGSATNEODSGCRS 339
 DB 56 RITERMAKAI---REMYCRECRERKDPRLTEIRYHKKCRERDGSRRASBPBEGGRKR 111
 QY 340 VGAD-----GT----- 345
 DB 112 PASDPELQRRASGTGVGAMLRGASAPRKSSPQPLVATPSQHNNHQQOQQOQIRRSAR 171
 QY 346 -----DC-----TSIGTV- 353
 DB 172 MCGECACRTEDGCHDCPCEMKKFGSPNKRQKCRRLRQCLRRARBSYKTPSSLSPT 231
 QY 354 -----EKSQ---EDQG-----IKGRIEKAAP----- 373
 DB 232 PSEALPRRRPPPTQQQPPQSQKLRIRNDEGTVLSVVKEPFATATPRLSDDLALD 291
 QY 374 -----SGKKKLKIFQ- 383
 DB 292 PDLVDPGAGAFDDHGLPMSDABSPFLDPLRRRAVKVKKRKKSEKKERRYKR 351
 QY 384 -----FVBAAPGPKICIGRCSSVAQPDSTYGSNDCLTKRAA 420
 DB 352 HRQKHNDKMKHPERADAKDPAFLPQCLGRCVRAAGPSKRYCSDDCMKLLAA 405
 RESULT 12
 ID CGBP_HUMAN STANDARD; PRT; 656 AA.
 AC Q9P0U4; Q96BC8; Q9P2V7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DB Cpg binding protein (Protein containing PHD finger and CXXC domain 1).
 GN CGBP OR PCCL1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20153771; PubMed=10688657;
 RA Shin Voo K., Carlone D.L., Jacobsen B.M., Flodin A., Skalniak D.G.;
 RT "Cloning of a mammalian transcriptional activator that binds
 unmethylated CpG motifs and shares a CXXC domain with DNA
 methyltransferase, human trithorax, and methyl-CpG binding domain
 protein 1.";
 RT Mol. Cell. Biol. 20:2108-2121(2000).
 RN [2]
 RP SEQUENCE FROM N.A., AND DNA-BINDING.
 RX MEDLINE=20261256; PubMed=10799292;
 RA Fujino T., Hasegawa M., Shibata S., Kishimoto T., Imai S.-I.,
 RA Takano T.;
 RT "PCCK1, a novel DNA-binding protein with PHD finger and CXXC domain,
 is regulated by proteolysis.";
 RT Biochem. Biophys. Res. Commun. 271:305-310(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=21154917; PubMed=11230166;
 RA Wiemann S., Weill B., Wellenreuther R., Gassenhuber J., Glasel S.,
 RA Ansgore W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
 RA Tauber J., Dueterhoeft A., Beyer A., Koehler K., Strack N.,
 RA Mewes H.-W., Oestermeider B., Obermaier B., Tampe J., Heubner D.,
 RA Wambutt R., Korn B., Klein M., Poustka A.;
 RT "Towards a catalog of human genes and proteins: sequencing and
 RT analysis of 500 novel complete protein coding human cDNAs.";
 RT Genome Res. 11:422-435(2001).
 RN [4]
 RP SEQUENCE FROM N.A.

RC TISSUE-Colon, and Skin;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Streusberg R. L., Feingold B. A., Grouse L. H., Derge J. G.,
 RA Klausner R. D., Collins P. S., Wagner L., Shemen C. M., Schuler G. D.,
 RA Altschul S. F., Zeeber B., Buetow K. H., Schaefer C. F., Bhat N. K.,
 RA Hopkins R. F., Jordan H., Moore T., Max S. I., Wang J., Hsieh F.,
 RA Datchenko L., Marusha K., Farmer A. A., Rubin G. M., Hong L.,
 RA Stapleton M., Soares M. B., Bonaldi M. F., Casavari T. L., Scheetz T. B.,
 RA Brownstein M. J., Ueda T. B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S. S., Loquellano N. A., Peters G. J., Abramson R. D., Mullaly S. J.,
 RA Bonak S. A., McEwan P. J., McKernan K. J., Malek J. A., Gunaratne P. H.,
 RA Richards S., Worley K. C., Hale S., Garcia A. M., Gay L. J., Hulyk S. W.,
 RA Villalón D. K., Kuzny D. M., Sodergren E. J., Lu X., Gibbs R. A.,
 RA Paley J., Helton B., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A. C., Shevchenko Y., Bouffard G. G.,
 RA Blakesley R. W., Touchman J. W., Green E. D., Dickson M. C.,
 RA Rodriguez A. C., Grimwood J., Schmutz J., Myers R. M.,
 RA Butterfield Y. S. N., Krzywinski M. J., Skalek U., Smalls D. E.,
 RA Scherch A., Schein J. E., Jones S. J. M., Marra M. A.,
 RA "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [5]
 RP DNA-BINDING DOMAIN.
 RX MEDLINE=21576202; PubMed=11572867;
 RA Lee J. H., Voo K. S., Skalniak D. G.,
 RT "Identification and characterization of the DNA binding domain of
 RT Cpg-binding protein.";
 RL J. Biol. Chem. 276:44669-44676(2001).
 CC -1- FUNCTION: Transcriptional activator that exhibits a unique DNA
 CC binding specificity for [AC]CpG[AC] unmethylated CpG motifs.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: Ubiquitous.
 CC -1- DOMAIN: The acidic domain carries the potential to activate
 CC transcription.
 CC -1- PTM: May be regulated by proteolysis.
 CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -1- SIMILARITY: Contains 1 CXXC-type zinc finger.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, A0149758; AAF37799.1; -
 DR EMBL, A0131063; BAA96307.1; -
 DR EMBL, AL131862; CAB66796.1; -
 DR EMBL, BC014940; AAH14940.1; -
 DR EMBL, BC015733; AAH15733.1; -
 DR GO: GO:0005634; C:nucleus; IDA.
 DR GO: GO:0016563; P:transcriptional activator activity; IDA.
 DR GO: GO:0006353; P:regulation of transcription, DNA-dependent; IDA.
 DR InterPro: IPR002857; Znf_CXXC.
 DR Pfam: PF00628; PHD; 1.
 DR Pfam: PF00628; Zf_CXXC; 1.
 DR SMART: SM00249; PHD; 1.
 DR PROSITE: PS01359; ZF_PHD_1; 1.
 DR PROSITE: PS50016; ZF_PHD_2; 1.
 KW Transcription regulation; Activator; DNA-binding; Zinc-finger; Zinc;
 KW Metal-binding; Coiled coil; Nuclear protein.
 FT ZN_FING 28 76
 FT ZN_FING 161 208
 FT ZN_FING 256 317
 FT DOMAIN 321 359
 FT DOMAIN 422 474
 FT DOMAIN 156 161
 FT DNA_BIND 162 221
 FT CONFLICT 117 117
 D -> N (IN RBP. 1).

FT CONFLICT 302 302 H -> N (IN RBP. 1).
 FT CONFLICT 340 340 K -> KWMER (IN REF. 4; AAH15733).
 SQ SEQUENCE 656 AA; 75711 MW; 6D2376B44905A18 CRC64;
 Query Match 5.2%; Score 166.5; DB 1; Length 656;
 Best Local Similarity 18.4%; Pred. No. 0.043;
 Matches 74; Conservative 31; Mismatches 55; Indels 243; Gaps 13;
 QY 245 EBNPBAKPKPKCEVYPNALYTCROPHNNRPMICRCCEMFRGDCVGSSEAGRL 304
 DB 15 EBSKSENGENAP-----IYCIKRPDINCFMIGCDNCNEMFAGDCIRLEKNAKAL 65
 QY 305 EBNGEDYICPNC---TILQVO-----DETNGSATNEODS----- 335
 DB 66 ---REWYCRERKEDPKLEIRYRKKSREKD---NENDSEPRDGGGRKRPVDDPDL 118
 QY 336 ---GCRSVGADGT-----TSICTV----- 353
 DB 119 QRAGSGTGVGAMLRGASPHKSSPQPLVATPSPQHQQOQOQIKRSAMGCEACRR 178
 QY 346 -DC----- 353
 DB 179 BDGHCDFCRDMKKCGPKIRKQCLRQOCLRARESYRFPSSLSPTPSSLPFRPR 238
 QY 354 ---EOKSG---BDQG---IKRIEKANP----- 373
 DB 239 LPTQOQPOPSQKLGIRBDEGAVASSTVKEPPEATPTPLDDEDLPLDPLYQDFCAGA 296
 QY 374 -----SGKKELKIFQ----- 383
 DB 299 PDHGLPMSDTEBSPFLDPALRKAVKGVYRREKSEKKEERYKRRQKHKDK 358
 QY 384 -----PVVEAPGAPKCTGPGCSSVAPQPSVYCSNDILKHA 420
 DB 359 KPERADADPASPPLQCLGPGCVRAPOPSKTCSPDCGMKLA 401
 DB
 RESULT 13
 DMP1 RAT STANDARD; PRT; 489 AA.
 AC P98193;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix
 DE protein-1) (DMP-1) (AG1).
 GN DMP1.
 OS Rattus norvegicus (Rat).
 OC Rattus norvegicus (Rat).
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Tooth;
 RX MEDLINE=93286101; PubMed=8509401;
 RA George A., Sabay B., Simmon P. A., Veis A.,
 RA "Characterization of a novel dentin matrix acidic phosphoprotein.
 RT Implications for induction of biomineralization.";
 RL J. Biol. Chem. 268:12624-12630(1993).
 CC -1- FUNCTION: Could be involved in the induction of mineralization of
 CC extracellular matrix.
 CC -1- TISSUE SPECIFICITY: Expressed in tooth particularly in odontoblast
 CC and ameloblast.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, L11354; -; NOT_ANNOTATED_CDS.

Biomineralization; Extracellular matrix; Signal.

FT SIGNAL 1 16 POTENTIAL.

FT CHAIN 17 489 DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.

FT DOMAIN 41 44 POLY-PRO.

FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SITE 334 336 CELL ATTACHMENT SITE (POTENTIAL).

SEQ SEQUENCE 489 AA; 53058 MW; 59P81479DDA085 CRC64;

Query Match 5.1%; Score 165.5; DB 1; Length 489;

Best Local Similarity 26.0%; Pred. No. 0.035;

Matches 72; Conservative 45; Mismatches 97; Indels 63; Gaps 12;

3 DKGH-----LSNEAPKAIKPTSK-----EPRKFWKRTTAKBAGADTEAD 46

205 DRGHTRMSSAGIRSEBSKGDHEPTSTQSDSDSQVEFSSRSFRSRVSSEDDRGEL-AD 263

47 PSEQQPOQHNLSTLRSGQPKRTVERVEFLTVRRRGRKNPVSLDESEPTSTVTDVE 106

264 SRSRTQSVSTEDFRS-KESRSSTQED--TAFTQSQEDSPSQDPSSSS-----E 312

107 TARGSVSSSRIRSGPVSDSLGKHPASSR-----KAKGBEE-----EDT 148

313 EAGEPSQSSSESGQVASESGNDPNTSGTGQRODSSESEEDPLNTFSSSSGSTEQQ 372

149 SDSDS-DGLTLKELQNLRRKREQPVERSLRSGQNLRLKK--RRREDSAFETGVSQISA 205

373 GDSSNSLSLSE-----ESQESAQDEDSGGLQSQASRSRSQESQSDRS 424

206 EDRPLCKQ-----EPASQGVVSQSTDDIENOLE 236

425 EENRDSQDSRSRKRBSNSTGTSSSEDDHPKNIE 461

RESULT 14

AK12_HUMAN STANDARD; PRT; 1781 AA.

AC Q02952; Q00310; Q00498; Q99970;

DT 01-FEB-1994 (Rel. 28, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE A-kinase anchor protein 12 (A-kinase anchor protein 250 kDa) (AKAP 250) (Myasthenia gravis autoantigen gravin).

DE AKAP12 OR AKAP250.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NCBI_TaxID=9606;

[1]

SEQUENCE FROM N.A. (ISOFORM 1).

RP TISSUE=Heart;

RC MEDLINE=97153077; PubMed=9000000;

RA Nauert J.B., Klauk T.M., Langeberg L.K., Scott J.D.;

RT "Gravin, an autoantigen recognized by serum from myasthenia gravis patients, is a kinase scaffold protein.";

RT Curr. Biol. 7:52-62(1997).

RL [2]

SEQUENCE FROM N.A. (ISOFORM 2).

RP TISSUE=Umbilical vein endothelial cells;

RC MEDLINE=98269042; PubMed=9604001;

RA Sato N., Kokame K., Shimokado K., Kato H., Miyata T.;

RT "Changes of gene expression by lysophosphatidylcholine in vascular endothelial cells: 12 up-regulated distinct genes including 5 cell growth-related, 3 thrombosis-related, and 4 others.";

RT J. Biochem. 123:1119-1126(1998).

RL [3]

SEQUENCE OF 43-1781 FROM N.A.

RP TISSUE=Umbilical vein endothelial cells;

RA Bowditch R.D., Ginsberg M.H.;

RT "Sequence of gravin cDNA isolated from a human umbilical vein endothelial cell library.";

RT Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.

[4]

RP SEQUENCE OF 1477-1781 FROM N.A.

RC TISSUE=Umbilical vein endothelial cells;

RK MEDLINE=92395179; PubMed=1522245;

RA Gordon T., Grove B., Loftus J.C., O'Toole T., McMillan R.,

RA Lindstrom J., Ginsberg M.H.;

RT "Molecular cloning and preliminary characterization of a novel cytoplasmic antigen recognized by myasthenia gravis sera.";

RT J. Clin. Invest. 90:992-999(1992).

RL [1]

FUNCTION: Anchoring protein that mediates the subcellular compartmentation of protein kinase A (PKA) and protein kinase C (PKC).

CC -1 SUBUNIT: Binds to dimeric RII-alpha regulatory subunit of PKC.

CC -1 SUBCELLULAR LOCATION: Cytoplasmic. May be part of the cortical cytoskeleton.

CC -1 ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=1;

CC IsoId=Q02952-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q02952-2; Sequence=VSP_004110, VSP_004111;

CC Note=No experimental confirmation available;

CC -1 TISSUE SPECIFICITY: Expressed in endothelial cells, cultured fibroblasts and osteosarcoma, but not in platelets, leukocytes, monocytic cell lines or peripheral blood cells.

CC -1 INDUCTION: Activated by lysophosphatidylcholine (lysoPC).

CC -1 DOMAIN: Polypeptide regions located between residues 265 and 556 are involved in binding PKC.

CC -1 DISEASE: ANTIBODIES TO THE C-TERMINAL OF GRAVIN CAN BE PRODUCED BY PATIENTS WITH MYASTHENIA GRAVIS (MG).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; U01607; AAC51366.1; -

CC EMBL; AF001504; AAB58938.1; -

CC EMBL; AB003476; BAA19927.1; -

CC EMBL; M96322; AAA35911.1; -

CC GeneW; HGNC:370; AKAP12.

DR MIM; 604698; -

DR GO; GO:0005737; C:cytoplasm; TAS.

DR GO; GO:0005079; F:protein kinase A anchoring activity; TAS.

DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; TAS.

DR InterPro; IPR001573; Pfam; anch.

DR Pfam; PF03832; Pfam; anch; 3.

DR Antigen; Repeat; Alternative splicing.

KW DOMAIN 603 633 AKAP 1.

FT DOMAIN 752 782 AKAP 2.

FT DOMAIN 797 827 AKAP 3.

FT DOMAIN 98 101 POLY-GLU.

FT DOMAIN 265 556 INVOLVED IN PKC-BINDING (PROBABLE).

FT DOMAIN 1540 1553 RII-BINDING (PROBABLE).

FT VARSPLIC 1 98 Missing (in isoform 2).

FT VARSPLIC 1 106 /FTId=VSP_004110.

FT VARSPLIC 99 106 REEIVTVR -> MGLTITIT (in isoform 2).

FT /FTId=VSP_004111.

FT CONFLICT 117 117 E -> K (IN REF. 2).

FT CONFLICT 142 144 NKN -> TPEI (IN REF. 2 AND 3).

FT CONFLICT 215 215 Q -> E (IN REF. 2).

FT CONFLICT 448 448 G -> E (IN REF. 2 AND 3).

FT CONFLICT 694 694 R -> G (IN REF. 2 AND 3).

FT CONFLICT 867 867 G -> S (IN REF. 2 AND 3).

FT CONFLICT 986 986 S -> A (IN REF. 2 AND 3).

FT CONFLICT 1530 1530 E -> E (IN REF. 3 AND 4).

FT CONFLICT 1581 1581 V -> M (IN REF. 4).

FT CONFLICT 1601 1601 Q -> L (IN REF. 2).

SEQ SEQUENCE 1781 AA; 191439 MW; BAA13937379PACOF CRC64;

Query Match 5.1%; Score 165.5; DB 1; Length 1781;
 Best Local Similarity 20.4%; Pred. No. 0.15;
 Matches 133; Conservative 81; Mismatches 201; Indels 237; Gaps 32;

QY 3 DKGLSN---EAPKA-----IKPTSKE-----FRKTW---GFRRT----- 32
 DB 136 DDGGEENRNIEQIPISSSNLELQPTESQANDIGFKVGFVGFVKDKDTEKPTV 195
 QY 33 ---TIAREGAGTBA---DPS-----EQOPOQHNLARSGRQPKRTREVERFL 76
 DB 196 QLLTVKXDBEGGAGADHPDPSLGAAPASKESEBPQ-----STKPEPTL 242
 QY 77 TTVARRKQAVPSLEDSSEPTSSVTVDVETASGVSSESEIRSGPVSDLSGK----- 130
 DB 243 KRBOSSHABISPPA---ESGQAVECKEKEGKEKPEKSKAESPTSETSGTFKKEFP 299
 QY 131 -----EHPASSKAKGEEBEDTSDSDGLTLKELQNLRRKREQPVERSLRSGQN 183
 DB 300 TQGMAGMRKKTSPFKPK---EDVEASB-----KKKEQGP----- 331
 QY 184 RLARKREED-SATGSGVQISAGQDRPLCKQEP-BASQGSVSGSETDDI-----ENQLEG 237
 DB 332 --EKVDTEBBDKAAVASEKLTASDAQHP---QEPASAHBEPLASRYEKVELPSEBQVSG 386
 QY 238 KATQANTENPRBAGKPKP-ECEVYDPVALYICROPHNNRPMICCDRCBEMFHQDCVGI 296
 DB 387 --SGSPSEB-----KPAPLATEVPDEKI-----EVQEEVVA-----EVHV 420
 QY 297 SEARGLLERNGBDYICPNCITLQVQDETNGSATNEODSGRSGVAGDGTCTSGTVEOK 356
 DB 421 STVEBETREBQKTE-----VEBTAGSVPAELVGMADAEPAQBARPAKELVLTKEET 468
 QY 357 --SGED--QGI-----KGRLEKANPSGKKKTKI-----PPVVEAPAPAKPI 395
 DB 469 CVSGEDPTQAGADLSPDEKYLKSPREGVASEVEMLSQGRMNVQSGPLKLTFTSTGLKLUS 528
 QY 396 G-----PGSSVAPQDPVYCSMDCLIKHAAA 421
 DB 529 GKQKQKRGCGDESGEHTQVPADSPSOEBOKGSSASSPBEPEITC-----LEKIDA 583
 QY 422 TMRPLSGKEQKTYKPEKVK-----TYPEKPSLEKCSVQVGIKISSVHKLASKEKNP 475
 DB 584 EVQDDGEABEGATSGEKKRGVTPWASFKQAVTPEKKV-----RRPSEBQKDB 633
 QY 476 VKXVWLASRSETSGKAAACESTPESKASDHNYNAVPEKPEKTLSPTLIS 527
 DB 634 LDYKXSATLST-----BSTASEMQEEMKGVSEBPKPEPKKRVDTSVS 677

RESULT 15
 UN89 CAEEL STANDARD; PRT; 6632 AA.
 AC 001761; Q17362;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
 GN UNC-89 OR C09D1.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodermidae; Caenorhabditis.
 OX NCB1_TaxID=6239;
 RN SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RP STRAIN=Bristol N2;
 RC MEDLINE=96180278; PubMed=8603916;
 RA Benlan G.M., Tinley T.L., Tang X., Borodovsky M.;
 RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
 RT assembly, encodes a giant modular protein composed of Ig and signal
 RT transduction domains";
 RL J. Cell Biol. 132:835-848 (1996).
 RP [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;
 RA Du Z., Le T.T., Wilson R.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP REVISIONS.
 RA Waterston R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Structural component of the muscle M-line. Myofibrillar
 CC lattice assembly begins with positional cues laid down in the
 CC basement membrane and muscle cell membrane. UNC-89 responds to
 CC these signals, localizes, and then participates in assembling an
 CC M-line.
 CC -1- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
 CC -1- SIMILARITY: Contains 1 DBL-homology (DB) domain.
 CC -1- SIMILARITY: Contains 1 fibronectin type III domain.
 CC -1- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 1 PH domain.
 CC -1- SIMILARITY: Contains 5 RCD domain.
 CC -1- SIMILARITY: Contains 1 SH3 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U33058; AB00542.1; --
 DR EMBL; AF003131; AB054132.2; --
 DR PDB; 1PHO; 20-DEC-00.
 DR WormPep; C09D1.1; CE30426.
 DR InterPro; IPR008957; FN III-like.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_C2.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR007850; RCD.
 DR InterPro; IPR000219; RhGEF.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00041; tm3; 1.
 DR Pfam; PF00047; Ig; 47.
 DR Pfam; PF00169; Ig; 1.
 DR Pfam; PF05177; RCD; 5.
 DR Pfam; PF00621; RhGEF; 1.
 DR Pfam; PF00018; SH3; 1.
 DR SMART; SM00408; IGC2; 23.
 DR SMART; SM00325; RhGEF; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS50010; DR 2; 1.
 DR PROSITE; PS50835; IG-LIKE; 49.
 DR PROSITE; PS50003; PH_DOMAIN; 1.
 DR PROSITE; PS50002; SH3; 1.
 KM Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
 KM 3D-structure.
 FT DOMAIN 63 127 SH3.
 FT DOMAIN 152 330 PH.
 FT DOMAIN 342 498 PH.
 FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
 FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.
 FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
 FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
 FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
 FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
 FT DOMAIN 1272 1315 THR-RICH.
 FT DOMAIN 1375 1475 RCD 1.
 FT DOMAIN 1479 1585 RCD 2.
 FT DOMAIN 1597 1695 RCD 3.
 FT DOMAIN 1700 1799 RCD 4.
 FT DOMAIN 1800 1860 RCD 5.
 FT DOMAIN 1862 2067 IG-LIKE C2-TYPE 7.
 FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comphen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2004, 21:19:55 ; Search time 48 Seconds
(without alignments)
3526.319 Million cell updates/sec

Title: US-09-787-016a-4
Perfect score: 3228
Sequence: 1 MDDKGLHSEBAPKAIKETS.....LKSQVLMKALSYSPFRPW 614

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1124875 seqs, 275673149 residues
Total number of hits satisfying chosen parameters: 1124875

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	740	22.9	181	9	US-09-925-301-1169
2	194.5	6.0	676	10	US-09-823-187-24
3	194.5	6.0	676	10	US-09-863-776-14
4	191.5	5.9	676	14	US-10-203-708-28
5	191.5	5.9	717	14	US-10-203-708-29
6	190.5	5.9	518	9	US-09-925-299-862
7	190.5	5.9	518	10	US-09-925-299-862
8	190.5	5.9	667	10	US-09-863-776-16
9	190.5	5.9	667	10	US-09-863-776-18
10	188.5	5.8	555	14	US-10-106-698-5847
11	188.5	5.8	670	10	US-09-823-187-86
12	188.5	5.8	670	10	US-09-863-776-51
13	188.5	5.8	671	10	US-09-946-374-308
14	188.5	5.8	671	10	US-09-823-187-85
15	188.5	5.8	671	10	US-09-863-776-20

16	188.5	5.8	671	12	US-10-206-915-346	Sequence 346, App
17	188.5	5.8	671	12	US-10-199-670-346	Sequence 346, App
18	188.5	5.8	671	12	US-10-201-858-346	Sequence 346, App
19	188.5	5.8	671	12	US-10-205-890-346	Sequence 346, App
20	188.5	5.8	671	12	US-10-208-024-346	Sequence 346, App
21	188.5	5.8	671	12	US-10-201-853-346	Sequence 346, App
22	188.5	5.8	671	12	US-10-174-581-346	Sequence 346, App
23	188.5	5.8	671	12	US-10-176-483-346	Sequence 346, App
24	188.5	5.8	671	12	US-10-176-749-346	Sequence 346, App
25	188.5	5.8	671	12	US-10-176-914-346	Sequence 346, App
26	188.5	5.8	671	12	US-10-176-915-346	Sequence 346, App
27	188.5	5.8	671	12	US-10-066-495A-308	Sequence 308, App
28	188.5	5.8	671	12	US-10-013-907A-308	Sequence 308, App
29	188.5	5.8	671	12	US-10-015-499A-308	Sequence 308, App
30	188.5	5.8	671	12	US-10-176-484-346	Sequence 346, App
31	188.5	5.8	671	12	US-10-180-550-346	Sequence 346, App
32	188.5	5.8	671	12	US-10-183-014-346	Sequence 346, App
33	188.5	5.8	671	12	US-10-187-728-346	Sequence 346, App
34	188.5	5.8	671	12	US-10-187-740-346	Sequence 346, App
35	188.5	5.8	671	12	US-10-187-883-346	Sequence 346, App
36	188.5	5.8	671	12	US-10-194-460-346	Sequence 346, App
37	188.5	5.8	671	12	US-10-194-463-346	Sequence 346, App
38	188.5	5.8	671	12	US-10-194-463-346	Sequence 346, App
39	188.5	5.8	671	12	US-10-195-884-346	Sequence 346, App
40	188.5	5.8	671	12	US-10-195-896-346	Sequence 346, App
41	188.5	5.8	671	12	US-10-196-744-346	Sequence 346, App
42	188.5	5.8	671	12	US-10-196-755-346	Sequence 346, App
43	188.5	5.8	671	12	US-10-196-757-346	Sequence 346, App
44	188.5	5.8	671	12	US-10-197-704-346	Sequence 346, App
45	188.5	5.8	671	12	US-10-197-704-346	Sequence 346, App

ALIGNMENTS

RESULT 1
US-09-925-301-1169
Sequence 1169, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: P4106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1169
LENGTH: 181
TYPE: PRT
ORGANISM: Homo sapiens
US-09-925-301-1169

Query Match	22.9%	Score 740;	DB 9;	Length 181;
Best local Similarity	65.1%	Pred. No. 3.2e-46;		
Matches 131;	Conservative 8;	Mismatches 15;	Indels 0;	Gaps 0;
QY	252	GKPKCEVYDPMALYICIGPNNPNNFMI	CCDRCEWPHQDVGISBARGRLIERNGEDY	311
DB	11	GRPKPCGSDYDPMALYICIGPNNPNNFMI	CCRCCEWPHQDVGISBARGRLIERNGEDY	70
QY	312	ICPNCITLQVODBTSHETADQSEARRPEDADCTDTSIGTIBQSSBDGIGRIEKKA		371
DB	71	ICPNCITLQVODBTSHETADQSEARRPEDADCTDTSIGTIBQSSBDGIGRIEKKA		130
QY	372	NPSGKKKLTIFQVYVYAPGAPKICIGPGCSSVAQP		405
DB	131	NPSGKKKLTIFQVYVYAPGAPKICIGPGCSSVAQP		164

```
RESULT 2
US-09-823-187-24
; Sequence 24, Application US/09823187
; Publication No. US2003096952A1
GENERAL INFORMATION:
; APPLICANT: Burgess, Catherine
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Liu, Xiaohong
; APPLICANT: Majumder, Kumud
; APPLICANT: Padigaru, Muralidhar
; APPLICANT: Patnajan, Meera
; APPLICANT: Shukets, Richard A
; APPLICANT: Spaderna, Steven K
; APPLICANT: Spytek, Kimberly
; APPLICANT: Taupier, Raymond J
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; PIR REFERENCE: 15966-745
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/193,339
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/193,205
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/195,343
PRIOR FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: 60/195,088
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/195,005
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/195,792
PRIOR FILING DATE: 2000-04-10
PRIOR APPLICATION NUMBER: 60/196,556
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: 60/197,081
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/197,525
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/197,087
PRIOR FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 103
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 24
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-823-187-24

Query Match      6.0%; Score 194.5; DB 10; Length 676;
Best Local Similarity 22.0%; Pred. No. 1,6e-05;
Matches 118; Conservative 83; Mismatches 183; Indels 153; Gaps 24;

QY 28 GFRRTT-----IAKREGADTEAD-----PSEOCFOQNNLSLRSGRQPKTERVERFL 76
DB 154 GUKRTPLAKVSVKRAKASDLDQASVPSFEENSSSSSEKTSDDPTPEK--KAA 211
QY 77 TTVRRRG-----KKNVPVSLDSSEPTSTVTVDVTASBGVSSESEIRSGFVSDSLK 130
DB 212 VAAPRRGPLGKKKKKAPASDSDSKADSDGAKPEVVAARASASSSSSSSSSDSDSVYK 271
QY 131 EHPASSE-----KAKGGBEED-----TSSDSDGL-TLKELONR-----LRR 167
DB 272 KPPRGKPAKBPUPKPRGGRKKPKPRPPSSSSSSSDSDSDDEVIRISEWRRRREARRLEARR 331
QY 168 KKEQEPVERSLRGSQNRLLKKGRREBDSATGVSQISG-----EODRPLCKQPEA- 218
DB 332 KKEQEBELRLRLEQKKEKERRRR--ADRGKRRSSGSSSGELAEDEBDFVKKRKRKR 389
QY 219 SGGPVQSSEETDDIENGLGKATQANTENPREAK--PRCEGVYDPAALYICICROPANN 276
DB 390 GRGPSSSSSBP-EABLE-----REAKKSAKPPQSSSTBP-----ARKP----- 427
QY 277 RRMICDRCEBPHGDCVGISPAKGRLLERNGEDYICPCTIIQVQDFTNGSATINQDSG 336
```

```
DB 428 -----GQKKEVR-----PEERQ 441
QY 337 CRSYGADGTCTSIGTWBQSGEODGIGRIEKAHPGKKTLL-----IFQPVYEA PG 390
DB 442 ARPVVERT-----RKRSBGFSMDRKVKKKKPEVEERLQJHSBIKALRVSDPD 492
QY 391 APKICGPGCSSVAQPDVYCSNDCLRH--AAATWRFLSSGKEOKTKPREKVKTKPEKFS 448
DB 493 VKRCL-----NALBEIGTQVTSQILQKNTDVATLK-----KIRRYKANKDVMKKAABVY 543
QY 449 LPKCSVQVGIKISVYK-RLASERKENPVKVMLASRSETSGKAAACSSSTPSWASD 504
DB 544 TRKSRVVGPKIEINVQYKNGMEKKAEBTL-----AGELAGEFLAGEBAPEKABD 597

RESULT 3
US-09-863-776-14
; Sequence 14, Application US/09863776
; Publication No. US20030198953A1
GENERAL INFORMATION:
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Majumder, Kumud
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Mishra, Vishnu
; APPLICANT: Padigaru, Muralidhar
; APPLICANT: Spaderna, Steven K
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Raateilil, Luca
; APPLICANT: Li, Li
; APPLICANT: Taupier, Raymond J
; APPLICANT: Gangolli, Bsha
; TITLE OF INVENTION: NO. US20030198953A1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-020
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: 09/540,763
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/206,679
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/206,688
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/206,829
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/207,748
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/207,798
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/208,263
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: 60/208,831
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: 60/209,451
PRIOR FILING DATE: 2000-06-05
PRIOR APPLICATION NUMBER: 60/210,060
PRIOR FILING DATE: 2000-06-07
PRIOR APPLICATION NUMBER: 60/219,507
PRIOR FILING DATE: 2000-07-20
PRIOR APPLICATION NUMBER: 60/221,337
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/221,927
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: 60/263,135
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: 60/263,688
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/263,694
PRIOR FILING DATE: 2001-01-24
NUMBER OF SEQ ID NOS: 155
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 14
; LENGTH: 676
; TYPE: PRT
```


Qy 131 EHPASSB-----KAKGGBEED-----TSDSDGL-TLKEIQR-----LRR 167
Db 313 KPRGRKPAKPLPGRGRKPRPSSSSSDSDSDVDRISWKRDDEARRELEARR 372
Qy 168 KRGEPVRSIRGSONRLRKRRREDSABTGSVQIGSA-----BODRPLCKQBPRA- 218
Db 373 RRGGBEELRLRLEQKKEKERRRR--ADRGAEKSGSGSSGDBLEBDEBPVKKGGRGR 430
Qy 219 SGPVSGSETDDIENQLEGKATQGTENPREAGK--PKPECEVYDPNALYICIQPHNN 276
Db 431 GRGPSSSDSRP-BAEILR-----REAKKSACKPOSSSTEP-----AKRP----- 468
Qy 277 RFLMICDRCEWPHDDCVGISRAGRLEBNGEDYICNCTILQVQDETNGSATNEQDSG 336
Db 469 -----GQEKERYR-----PBEKQ 482
Qy 337 CRVYAGDGTCTSIGTVBOKSGBDGIGRIEKAIPSGKKLK-----IFQVYBAAG 390
Db 483 AKPVYVERT-----RKSGSGSMRKRYEKKKPSVBERLQKLSHSEIKFALKVDSBD 533
Qy 391 APKCIGPCSSVAQPDVYCSNDCTIKR--AAATRPPLSSGKQKTKPKREKVKYKPEKFS 448
Db 534 VKRCL-----NMLEBGLQVTSQILQKNTDVVATLK-----KIRRYKANKDWMERKAAYV 584
Qy 449 LPKGVQVIGITSSVHK--RLASERKENPVKVMLASRSETSGKEAACSSSTPSASD 504
Db 585 TRLSRVLPPIEAVQVYKAKMEKKAEBKL--AGEBLAGEBLAGEBAPOKKAED 638

RESULT 6

US-09-925-299-862
; Sequence 862, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 862
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (476)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-862

Query Match 5.9%; Score 190.5; DB 9; Length 518;

Best Local Similarity 22.0%; Pred. No. 2.2e-05;
Matches 119; Conservative 80; Mismatches 185; Indels 157; Gaps 25;

Qy 28 GPRRT-----IARBGAGDTEAD-----PSEQPOQHNLRLRSGROPKTERVEBFL 76
Db 38 GIKRTPLAKMSVSRARAKASDLDQASVSPERENSSSSSKRTSDDPFPER--KXA 95
Qy 77 TTVRRG-----KNQVPSLEDSSEPTSTVTDVETASBGVSSSEIRSGPVSDSLCK 130
Db 96 VRAPRRGLGGRKKKAAVSASDSKADSDAKPRVAMARASASSSSSSSDSDVSVK 155
Qy 131 EHPASSB-----KAKGGBEED-----TSDSDGL-TLKEIQR-----LRR 167
Db 156 KPRGRKPAKPLPGRGRKPRPSSSSSDSDSDVDRISWKRDDEARRELEARR 215
Qy 168 KRGEPVRSIRGSONRLRKRRREDSABTGSVQIGSA-----BODRPLCKQBPRA- 218

Db 216 RRGGBEELRLRLEQKKEKERRRR--ADRGAEKSGSGSSGDBLEBDEBPVKKGGRGR 273
Qy 219 SGPVSGSETDDIENQLEGKATQGTENPREAGK--PKPECEVYDPNALYICIQPHNN 276
Db 274 GRGPSSSDSRP-BAEILR-----REAKKSACKPOSSSTEP-----AKRP----- 311
Qy 277 RFLMICDRCEWPHDDCVGISRAGRLEBNGEDYICNCTILQVQDETNGSATNEQDSG 336
Db 312 -----GQEKERYR-----PBEKQ 325
Qy 337 CRVYAGDGTCTSIGTVBOKSGBDGIGRIEKAIPSGKKLK-----IFQVYBAAG 390
Db 326 AKPVYVERT-----RKSGSGSMRKRYEKKKPSVBERLQKLSHSEIKFALKVDSBD 376
Qy 391 APKCIGPCSSVAQPDVYCSNDCTIKR--AAATRPPLSSGKQKTKPKREKVKYKPEKFS 448
Db 428 TRLSRVLPPIEAVQVYKAKMEKKAEBKLAGEBLAGEBAPOKKGQAQHRSLSPSE 487
Qy 501 W 501
Db 488 W 488

RESULT 7

US-09-925-299-862
; Sequence 862, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 862
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (476)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-862

Query Match 5.9%; Score 190.5; DB 10; Length 518;

Best Local Similarity 22.0%; Pred. No. 2.2e-05;
Matches 119; Conservative 80; Mismatches 185; Indels 157; Gaps 25;

Qy 28 GPRRT-----IARBGAGDTEAD-----PSEQPOQHNLRLRSGROPKTERVEBFL 76
Db 38 GIKRTPLAKMSVSRARAKASDLDQASVSPERENSSSSSKRTSDDPFPER--KXA 95
Qy 77 TTVRRG-----KNQVPSLEDSSEPTSTVTDVETASBGVSSSEIRSGPVSDSLCK 130
Db 96 VRAPRRGLGGRKKKAAVSASDSKADSDAKPRVAMARASASSSSSSSDSDVSVK 155
Qy 131 EHPASSB-----KAKGGBEED-----TSDSDGL-TLKEIQR-----LRR 167
Db 156 KPRGRKPAKPLPGRGRKPRPSSSSSDSDSDVDRISWKRDDEARRELEARR 215
Qy 168 KRGEPVRSIRGSONRLRKRRREDSABTGSVQIGSA-----BODRPLCKQBPRA- 218
Db 216 RRGGBEELRLRLEQKKEKERRRR--ADRGAEKSGSGSSGDBLEBDEBPVKKGGRGR 273


```

CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: 09/540,763
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/206,679
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/206,688
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/206,829
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/207,748
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/207,798
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/208,263
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: 60/208,831
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: 60/209,451
PRIOR FILING DATE: 2000-06-05
PRIOR APPLICATION NUMBER: 60/210,060
PRIOR FILING DATE: 2000-06-07
PRIOR APPLICATION NUMBER: 60/219,507
PRIOR FILING DATE: 2000-07-20
PRIOR APPLICATION NUMBER: 60/221,337
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/221,927
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: 60/263,135
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: 60/263,688
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/263,694
PRIOR FILING DATE: 2001-01-24
NUMBER OF SEQ ID NOS: 155
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 18
LENGTH: 667
TYPE: PRT
ORGANISM: Homo sapiens
US-09-863-776-18

```

```

Query Match      5.9%: Score 190.5; DB 10; Length 667;
Best Local Similarity 22.5%; Pred. No. 3.1e-05;
Matches 120; Conservative 84; Mismatches 179; Indels 153; Gaps 25;

```

```

28 GFRRTT-----IARREGAGDTAD-----PSEQOPQOHNLSLRSGRPKRTERVEEFL 76
154 GLKRTPLALMSVSKARARASDLDOASVSPSEENSSSESEKTSDDPTPEK--KAA 211
77 TTVRRRG-----KKNVPVSLDSSEPTSTVTVDVETASGVSSESEIRSGPVSLSLK 130
212 VVAPRRGPGKGRKKKAPASDSKADSDGAKPEFVAARASASSSSSSSSSDSVSYK 271
131 EHPASSE-----KAKGSEBED-----TSDSDSLG-TLKELONR-----LRR 167
272 KPRGRKPAKPKLPKPRGRKPKPRPPSSSSSDSDSDDEVDRISSEWRROBARRELEARR 331
168 KROEPVERSLGSGONRLKKRREBDSATGVSQISA-----EODRPLCKQPEA--SOG 221
332 RROQEBELRLRL-BOHEKERRRRRAARSGS--GSSGDELEBDEBPVYKGRKGRGK 388
222 PVSQSEITDIDLEQLEKATQGTENPREAGK--PRCEBYVDPNALYICICROPHNNRM 279
369 PPSSSDSEB-BALEB-----REAKKSARKPQSSSTP-----ARKP-----423
260 ICCDCEWPHGDCVGISBARGLLERNGEDYICPNTLLIQVODETNGSATVQDSGCRS 339
424 -----GQKEKVR-----PEKQQAKP 440
340 VVADGDTCTSIGTVEQKSGEDGIGRIEYKANPSGKKLK-----IFQPVVEAPGAKR 393
441 VVVERT-----RKRSBGFSMDRKVKKKPEPSVEKLOKLHSEIKALKYDSIDVR 491

```

```

394 CIGPCSSVAOPDSVYCSNDCLIKH--AAATRFISGCKEQTKREKVKTPKPSLJK 451
492 CL-----NALLEGTQVTSQILQNTDVATLK-----KIRRYKANKVMEGAAYTTL 542
452 CSVQGIKISSYVK-RLASKEKENPVKKVMLASRSETSGKEAACSSSTPMSAD 504
543 KSRVIGPTEAVQKVDKAGMEKEKAEKL-----AGBELAGBAPQKABDPKST 593

```

```

RESULT 10
US-10-106-698-5847

```

```

Sequence 5847, Application US/10106698
Publication No. US20030109690A1

```

```

GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
FILE REFERENCE: PA005P1
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 5847
LENGTH: 555
TYPE: PRT
ORGANISM: Homo sapiens
US-10-106-698-5847

```

```

Query Match      5.8%: Score 188.5; DB 14; Length 555;
Best Local Similarity 21.6%; Pred. No. 3.3e-05;
Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;

```

```

28 GFRRTT-----IARREGAGDTAD-----PSEQOPQOHNLSLRSGRPKRTERVEEFL 76
38 GLKRTPLALMSVSKARARASDLDOASVSPSEENSSSESEKTSDDPTPEK--KAA 95
77 TTVRRRG-----KKNVPVSLDSSEPTSTVTVDVETASGVSSESEIRSGPVSLSLK 130
96 VVAPRRGPGKGRKKKAPASDSKADSDGAKPEFVAARASASSSSSSSDSVSYK 155
131 EHPASSE-----KAKGSEBED-----TSDSDSLG-TLKELONR-----LRR 167
156 KPRGRKPAKPKLPKPRGRKPKPRPPSSSSSDSDSDDEVDRISSEWRROBARRELEARR 215
168 KROEPVERSLGSGONRLKKRREBDSATGVSQISA-----EODRPLCKQPEA--218
216 RROQEBELRLRLBOHEKERRRR-ADRGEARSGSSGDELEBDEBPVYKGRKGRK 273
219 SGPVQSEITDIDLEQLEKATQGTENPREAGK--PRCEBYVDPNALYICICROPHNN 276
274 GRGPSSDSEB-BALEB-----REAKKSARKPQSSSTP-----ARKP-----311
277 RRMICCDCEWPHGDCVGISBARGLLERNGEDYICPNTLLIQVODETNGSATVQDSG 336
312 -----GQKEKVR-----PEKQQ 325
337 CRVAGADGDTCTSIGTVEQKSGEDGIGRIEYKANPSGKKLK-----IFQPVVEAPG 390
326 AKPVKVERT-----RKRSBGFSMDRKVKKKPEPSVEKLOKLHSEIKALKYDSID 376
391 APKCTIGPGSSVAOPDSVYCSNDCLIKH--AAATRFISGCKEQTKREKVKTPKPSLJK 448
377 VKRCL-----NALLEGTQVTSQILQNTDVATLK-----KIRRYKANKVMEGAAY 427
449 LKRCISVQGIKISSYVK-RLASKEKENPVKKVMLASRSETSGKEAACSSSTPMSAD 504
428 TLKRSVIGAPKLEAVQKVDKAGMEKEKAEKL-----AGBELAGBAPQKABDPKST 481

```

RESULT 11

US-09-823-187-86
Sequence 86, Application US/09823187
Publication No. US20030096952A1
GENERAL INFORMATION:
APPLICANT: Burgess, Catherine
APPLICANT: Gusev, Vladimir Y
APPLICANT: Liu, Xiaohong
APPLICANT: Majumder, Kumud
APPLICANT: Padigar, Muralidhar
APPLICANT: Patnirajan, Meera
APPLICANT: Shimeles, Richard A
APPLICANT: Spaderna, Steven K
APPLICANT: Spytek, Kimberly
APPLICANT: Taupier, Raymond J
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 15966-745
CURRENT APPLICATION NUMBER: US/09/823,187
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/193,339
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/193,205
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/195,343
PRIOR FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: 60/195,088
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/195,005
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/195,792
PRIOR FILING DATE: 2000-04-10
PRIOR APPLICATION NUMBER: 60/196,556
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: 60/197,081
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/197,525
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/197,087
PRIOR FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 103
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 86
LENGTH: 670
TYPE: PRT
ORGANISM: Homo sapiens
US-09-823-187-86

Query Match 5.8%; Score 188.5; DB 10; Length 670;
Best Local Similarity 21.6%; Pred. No. 4.3e-05;
Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;

28 GRRRTT-----IAREGAGDTREAD-----PSEQOQOHNLSLRSGRQPKTTERREBEL 76
154 GIKRKTIPALMWSVSKARASDLDQASVSPSEKSSSSSSBKTSDDPTERK--KAA 211
77 TTVRRG-----KKNVPSLEDSSEPTSTVDTVYASGVSSEIRSGFVPSLIG 130
212 VRAPRGPIAGRGKKAAPSASDSSTADSDGAKPBPVMAARSSSSSSSSSSSDSDSVK 271
131 BHPASB-----KAKGHEBED-----TSDSDSDGL-TLKEIQNR-----LRR 167
272 KPRGRKPALEKPLPKRGKRGKPKRPPSSSSSSSDSDSDVDRISEWKRDRARBLEARR 331
168 KBOEPVERSLKSGQNLKRGKREBDSATGVSQVIGSA-----EQRPPLCKQEPBA 218
332 RREOEBELRLRLEQBEKREKRRER--ADRGAEKSSGSSGDELRRDDEPVKRGKGR 399
219 SGGVPSQSHTDIENLEGATQANTENRBAK--PRPECEVDPNALYCICROPHN 276
390 GNGPSSSSSEP-EAELF-----RAKKAKEKQSSSTEP-----AKRP--- 427
277 RMVICCDRCBWFHGDVGISBARGLLBRNGEDYICPNCTIIQVDETNGSATNEQDSG 336

RESULT 12

US-09-863-776-51
Sequence 51, Application US/09863776
Publication No. US20030198953A1
GENERAL INFORMATION:
APPLICANT: Spytek, Kimberly A
APPLICANT: Majumder, Kumud
APPLICANT: Tchernev, Velizar T
APPLICANT: Mishra, Vishnu
APPLICANT: Padigar, Muralidhara
APPLICANT: Spaderna, Steven K
APPLICANT: Shenoy, Suresh G
APPLICANT: Rastelli, Luca
APPLICANT: Li, Li
APPLICANT: Taupier, Raymond J
APPLICANT: Gangoli, Saba
TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-020
CURRENT APPLICATION NUMBER: US/09/863,776
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: 09/540,763
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/206,679
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/206,688
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/206,829
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/207,748
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/207,798
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/208,263
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: 60/208,831
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: 60/209,451
PRIOR FILING DATE: 2000-06-05
PRIOR APPLICATION NUMBER: 60/210,060
PRIOR FILING DATE: 2000-06-07
PRIOR APPLICATION NUMBER: 60/219,507
PRIOR FILING DATE: 2000-07-20
PRIOR APPLICATION NUMBER: 60/221,337
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/221,927
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: 60/263,135
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: 60/263,688
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/263,694
PRIOR FILING DATE: 2001-01-24
NUMBER OF SEQ ID NOS: 155
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 51
LENGTH: 670
TYPE: PRT

428 -----GQEKRYR-----PEEKQ 441
337 CSRVGADGTCTSIGTVBQSGEDQGIKRIEKAANPSGKKLK-----IFQPVYAPG 330
442 ADFVVERI-----RKSGEFSMDRKVEKKEGFSVEKGLQKHSEIKFALKTDSIDP 492
391 AFKCIQGCCSSVAGPDSVYCSNDCLIKH--AAATRFPLSSGKQKTPKPKXVTKREKTS 448
493 VRCL-----NALREIGTLQVTSQILQKNTDVVAATK-----KIRRYKANDVNEKAAEVY 543
449 LKCSQVQGIKISVYK-RLASEKRENPKVMASRSETSKGAAACSSSTPSMASD 504
544 TELKSRVLSPKTEIAQVKNKAKGEKEKAEKL---AGEBLAGEBALQEKAEADPSTD 597

ORGANISM: Homo sapiens
US-09-863-776-51

Query Match 5.8%; Score 188.5; DB 10; Length 670;
Best Local Similarity 21.6%; Pred. No. 4.3e-05;
Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;

```
QY 28 GRRRTT-----IANKRGAGDTGAD-----PSEQCPQOQHNLISLRSGRQPKTERVERBL 76
DB 154 GIKRTPALNKMVSFKARAKASDLDQASVSPSESESSSSSESEKTSDDPTEPK--KAA 211
QY 77 TVVRRG-----KNVPSLEDSSEPTSTVTDVETASRGVSSEBSIRSGVPSDLCK 130
DB 212 VAPRRGLGRKKKAPASDSADSDGAKPEPVANARSSSSSSSSSSSSSDSDSVYK 271
QY 131 EHPASSR-----VAKGEBEED-----TSDSDSGCL-TLKEQLNR-----LRR 167
DB 272 KPRGRKPAKPLPKPRGRKKPKPRPSSSSSDSDSDVDRISEWKRREDEARRRELEARR 331
QY 168 KKEQEPVERSLRGSQNRLLKKRREDSATGVSVOIGSA-----EODRPLCKQPEA- 218
DB 332 RREQEBELRLRBOEKERERRR--ADRGELRSGSGSGDELREDEPVEKRGKGR 389
QY 219 SGGPSQSRTDIDENLBKATQATNEENPREAKG--PKPCEVYDPMALYICRQPHNN 276
DB 390 GGGPSSSDSEP-EKELR-----REKKSAKKPQSSSTP-----AKP----- 427
QY 277 RFMICCDRCCEWPHGCVGISBARGRLRLRNGBDYICPCTLLQVQDETNGSATNEDQSG 336
DB 428 -----GQKEKRVX-----PREKQ 441
QY 337 CRVAGADGDTCTSIGTVEOKSGEDQIKRIEKAAPSGKTKL-----IQPVYEAAG 390
DB 442 ARPVYKERT-----RKRSQFSGMDRKVEKKKPPVEEKLQKXLSKPKALKVDSPD 492
QY 391 APKCIQPCSSVAQPSVYCSNDCLKH--AAATNPLFSGKQKTKPKKVKTKDEKS 448
DB 493 VKRCL-----NALBELGTLOVTSQILOKNDVYATLK-----KIRKANKDWEKAAAEVY 543
QY 449 LPKCSVQVGKIKISVYHK-RLASREKREPVKQWMLASRSTSGKAAECSTSPSWASD 504
DB 544 TLKKSIVLGPXTLEAVQKVKAKGHEKKAEBKL--AGBELAGBEAQEKABEDKPSFD 597
```

RESULT 13

US-09-946-374-308

Sequence 308, Application US/09946374

Publication No. US20030073129A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Baton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paonl, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumes, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
ACIDS
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2830P1C1

CURRENT APPLICATION NUMBER: US/09/946,374
CURRENT FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: 60/098716
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098723
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098749
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098750
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098803
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098821
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098843
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/099536
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099602
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099642
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099741
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099754
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099763
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099792
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099808
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099815
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099816
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100385
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100388
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100390
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100584
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100627
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100661
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100662
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100664
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100683
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100684
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100710
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100711
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100848
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100849
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100919
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100930

PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101014
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101068
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101071
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101279
PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: 60/101471
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101472
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101474
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101475
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101476
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101477
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101479
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101741
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101743
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101915
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101916
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/102207
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102240
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102307
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102330
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102331
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102484
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102487
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102570
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102571
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102684
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102687
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102965
PRIOR FILING DATE: 1998-10-02
PRIOR APPLICATION NUMBER: 60/103258
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/103314
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103315
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103328
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103395
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103396
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103401
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103449
PRIOR FILING DATE: 1998-10-06

PRIOR APPLICATION NUMBER: 60/103633
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103678
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103679
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103711
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/104257
PRIOR FILING DATE: 1998-10-14
PRIOR APPLICATION NUMBER: 60/104987
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105000
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105002
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105104
PRIOR FILING DATE: 1998-10-21
PRIOR APPLICATION NUMBER: 60/105169
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105266
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105693
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105694
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105807

Query Match 5.8%; Score 188.5; DB 10; Length 671;
Best Local Similarity 21.6%; Pred. No. 4,3e-05;
Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;

QY 28 GPRRTT-----IAKRGAGDTAD-----PSEQPOQHNLSRSGROPKTERVEEFL 76
DB 154 GLKRTPLAKMSVSRKAKASDLDQASVSPSESESESSSEKSDODFTPEK--KAA 211
QY 77 TTVRRRG-----KQNPVLSLDSSEPTSTYTDVETASGVSSESEIRSGPVSDSLK 130
DB 212 VRAPRRGLGRRKKKAPASDSDSKAUSDGAKPEVAMARASSSSSSSSDSDSVSK 211
QY 131 BHPASSE-----YAKGGBREED-----TSDSDSDGL-TLKEIQNR-----LNR 167
DB 272 XPPRGKRPAREKLPKPRKPKPRPPSSSSSDSDSDVDRISSEWRBDEARRELEARR 331
QY 168 KREQBPVERSLRGSGNRLKKRREBDSATGVSVOIGSA-----EQRPLCKQPEA- 218
DB 332 RRDQEEELRLRQEKREKRRRR--ADRGALRSGSGSSGDELREDDBPVKRGRKGR 389
QY 219 SGQPVQSSTDIENOLEKATQGTENPREAGK--PQDECVVDPNMLYCIQROPHN 276
DB 390 GRGPPSSSDSRP-EARLE-----RRAKSAKTPQSSSTP-----ARKP--- 427
QY 277 RFNWCDCRCEBWFHGDVCSISBAGRLERNGEDYICPNCTLLQVODETNGSATNEQDSC 336
DB 428 -----GQKRYR-----PEKQ 441
QY 337 CSBVGADGTDCTSIGTVQKSGEDGIGRIKRYANPSEKTKLK-----TFQPVYBAP 390
DB 442 AKPVKYERT-----RKRSBGFMDRKVEKKEPSVEKLOKLSHSEIKFALKYDSDPD 492
QY 391 APKCIQPGCSSVAQDPVSVCNSDCLIKH--AAATAPFLSSGQKTKPKXKVTIKPEKFS 448
DB 493 VRKCL-----NALRELTQVTSQILQKNTDVATLK-----KIRRYKANDVKEKAEVY 543
QY 449 LPKCSVOVGIKISVYK--RLASBKRENPKVWTLASRSETSGKAAECSTPSPASD 504
DB 544 TLKSRVLDPKIYAVQKYNKAKGKREKABKRL--AGEELAGBRAQKAEADKPSD 597

RESULT 14
US-09-823-187-85
; Sequence 85, Application US/09823187
; Publication No. US20030096952A1

GENERAL INFORMATION:
APPLICANT: Burgess, Catherine
APPLICANT: Gusev, Vladimir Y
APPLICANT: Liu, Xiaohong
APPLICANT: Majumder, Kumud
APPLICANT: Padigar, Muralidhar
APPLICANT: Patturajan, Meera
APPLICANT: Shinkens, Richard A
APPLICANT: Spaderna, Steven K
APPLICANT: Spytek, Kimberly
APPLICANT: Taupier, Raymond J
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 15966-745
CURRENT APPLICATION NUMBER: US/09/823,187
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/193,339
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/193,205
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/195,343
PRIOR FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: 60/195,088
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/195,005
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/195,792
PRIOR FILING DATE: 2000-04-10
PRIOR APPLICATION NUMBER: 60/196,556
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: 60/197,081
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/197,525
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/197,087
PRIOR FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 103
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 85
LENGTH: 671
TYPE: PRT
ORGANISM: Homo sapiens
US-09-823-187-85

Query Match 5.8%; Score 188.5; DB 10; Length 671;

Best Local Similarity 21.6%; Pred. No. 4.3e-05; Mismatches 182; Indels 153; Gaps 24;

Matches 116; Conservative 86; Mismatches 182; Indels 153; Gaps 24;

QY 28 GFRATT-----IAKREGAGDTEAD-----PSEOPFOCHNLILRRSGRQPKTERVEREPL 76
DB 154 GUKRTPLAKMSVSKARAKASDLDQASVSEERNSSESSSEKTSDDPTPER--KAA 211
QY 77 TTVRRRG-----KKNVPVSLDSSEPTSTVTVDVETASGEGVSSSEIRSGPVSDSLAK 130
DB 212 VRAPRGPGKGRKKKAPASDSKADSDAKPEPVAMARASSSSSSSSDSDSVSK 271
QY 111 EHPASSF-----KAKGSEED-----TSDSDGL--TKELQNR-----LBR 167
DB 272 KPRGRKPAEKPLPFRGKPKPRPPSSSSSDSDSDVDRISEWGRREARRELEAKR 331
QY 168 KKEQFVRSLLSGONRLAKKRREDSAVTGSVQIGSA-----EODRPLCKQEPFA- 218
DB 332 RREGEELRLRLBQKEKERRRR--ADRGAEKNGSGSSSDDELREDEPVRKXGRKR 389
QY 219 SOGPVQSSETDDIENQLEGKATQNTENPREAGK--PRGCEVYDPAALYICICROPHNN 276
DB 330 GRGPSSSSSEF--EAELE-----REAKTSAKRPOSSSTEP-----AKRP-- 427
QY 277 RRMICDRCBEMFHDDCVGISARGLLEARNEDYICPCTTLOVDGDTNGSATNEDGSG 336
DB 428 -----GQKEKRVF-----PEKKQ 441
QY 337 CSRVGADTDCSTIGTVEQSGEDGIGKRIEKAANPSGKKLK-----IFQPVVEAPG 390
DB 337 CSRVGADTDCSTIGTVEQSGEDGIGKRIEKAANPSGKKLK-----IFQPVVEAPG 390

DB 442 AKPVKERT-----RKSEGPSMDRKVEKKEPSVEEKLQKLHSEIKELKALVDSBD 492
QY 391 APKCIIGPCSSVAOPDSVYCNIDCLKH--AAATRFISGGEOKTKPKREKPEKS 448
DB 493 VKRCL-----NMLEEGTLQVTSQILQKNTDVATLK-----KIRYKANKDVMEKAAEY 543
QY 449 LPKCSVQYIKISVHK--RLASERENPVKVMLASRSETSGKEAACSTPSMASD 504
DB 544 TRKSRVIGPTEIENQYKTKGMEKKEKAEKL--AGSELAGEBAPOEVAIDKPSFD 597

RESULT 15

US-09-863-776-20

Sequence 20, Application US/09863776

Publication No. US20030198953A1

GENERAL INFORMATION:

APPLICANT: Spytek, Kimberly A

APPLICANT: Majumder, Kumud

APPLICANT: Tchernev, Velizar T

APPLICANT: Mishra, Vishnu

APPLICANT: Padigar, Muralidhar

APPLICANT: Spaderna, Steven K

APPLICANT: Shenoy, Suresh G

APPLICANT: Rastelli, Luca

APPLICANT: Li, Li

APPLICANT: Taupier, Raymond J

APPLICANT: Gangolli, Esba

TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 21402-020

CURRENT APPLICATION NUMBER: US/09/863,776

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: 09/540,763

PRIOR FILING DATE: 2000-03-30

PRIOR APPLICATION NUMBER: 60/206,679

PRIOR FILING DATE: 2000-05-24

PRIOR APPLICATION NUMBER: 60/206,688

PRIOR FILING DATE: 2000-05-24

PRIOR APPLICATION NUMBER: 60/206,829

PRIOR FILING DATE: 2000-05-24

PRIOR APPLICATION NUMBER: 60/207,748

PRIOR FILING DATE: 2000-05-30

PRIOR APPLICATION NUMBER: 60/207,798

PRIOR FILING DATE: 2000-05-30

PRIOR APPLICATION NUMBER: 60/208,263

PRIOR FILING DATE: 2000-05-31

PRIOR APPLICATION NUMBER: 60/208,831

PRIOR FILING DATE: 2000-06-02

PRIOR APPLICATION NUMBER: 60/209,451

PRIOR FILING DATE: 2000-06-05

PRIOR APPLICATION NUMBER: 60/210,060

PRIOR FILING DATE: 2000-06-07

PRIOR APPLICATION NUMBER: 60/219,507

PRIOR FILING DATE: 2000-07-20

PRIOR APPLICATION NUMBER: 60/221,337

PRIOR FILING DATE: 2000-07-26

PRIOR APPLICATION NUMBER: 60/221,927

PRIOR FILING DATE: 2000-07-31

PRIOR APPLICATION NUMBER: 60/263,135

PRIOR FILING DATE: 2001-01-19

PRIOR APPLICATION NUMBER: 60/263,688

PRIOR FILING DATE: 2001-01-24

PRIOR APPLICATION NUMBER: 60/263,694

PRIOR FILING DATE: 2001-01-24

NUMBER OF SEQ ID NOS: 155

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO: 20

LENGTH: 671

TYPE: PRT

ORGANISM: Homo sapiens

US-09-863-776-20

Query Match 5.8%; Score 188.5; DB 10; Length 671;

Best Local Similarity 21.6%; Pred. No. 4.3e-05;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2004, 21:17:09 ; Search time 23 Seconds
(without alignments)
1378.189 Million cell updates/sec

Title: US-09-787-016a-4
Perfect score: 3228
Sequence: 1 MDDGHLNBRAPKIKRFTS.....LKSCVGLMLRAISYRFRPW 614

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCITUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	197.5	6.1	1805	1	US-07-853-913-2 Sequence 2, Appl
2	166	5.1	688	3	US-09-141-047-8 Sequence 8, Appl
3	165.5	5.1	1780	1	US-08-769-309A-5 Sequence 5, Appl
4	165.5	5.1	1780	3	US-08-994-570-5 Sequence 5, Appl
5	164	5.1	238	4	US-09-257-179-80 Sequence 80, Appl
6	153.5	4.8	1848	4	US-08-296-791-6 Sequence 6, Appl
7	153.5	4.8	1848	4	US-09-839-996-6 Sequence 6, Appl
8	153.5	4.8	1848	4	US-10-080-505-6 Sequence 6, Appl
9	153.5	4.8	1848	5	PCT-US95-10651A-6 Sequence 6, Appl
10	151.5	4.7	532	1	US-08-285-440-5 Sequence 5, Appl
11	151.5	4.7	532	1	US-08-630-349-5 Sequence 5, Appl
12	149.5	4.6	2468	4	US-09-976-594-726 Sequence 726, App
13	148	4.6	657	3	US-08-893-852A-3 Sequence 3, Appl
14	148	4.6	657	3	US-08-821-818-3 Sequence 3, Appl
15	148	4.6	657	4	US-09-052-753B-3 Sequence 3, Appl
16	144.5	4.5	1088	4	US-09-130-242-2 Sequence 2, Appl
17	142	4.4	1596	4	US-08-978-277A-4 Sequence 4, Appl
18	141.5	4.4	1739	4	US-09-976-594-76 Sequence 76, Appl
19	141	4.4	8991	4	US-08-714-741-32 Sequence 32, Appl
20	140.5	4.4	558	1	US-08-285-440-6 Sequence 6, Appl
21	140.5	4.4	558	1	US-08-630-349-6 Sequence 6, Appl
22	138.5	4.3	2289	3	US-09-051-019-2 Sequence 2, Appl
23	135.5	4.2	1233	4	US-09-645-456A-35 Sequence 35, Appl
24	135.5	4.2	1233	4	US-09-425-324A-35 Sequence 35, Appl
25	135.5	4.2	1233	4	US-09-645-791-35 Sequence 35, Appl
26	134.5	4.2	1231	4	US-09-595-684B-23 Sequence 23, Appl
27	134.5	4.2	1297	4	US-09-688-188B-14 Sequence 14, Appl

28	134.5	4.2	1297	4	US-09-291-417D-14	Sequence 14, Appl
29	134.5	4.2	1332	4	US-09-645-456A-9	Sequence 9, Appl
30	134.5	4.2	1332	4	US-09-425-324A-9	Sequence 9, Appl
31	134.5	4.2	1332	4	US-09-645-791-9	Sequence 9, Appl
32	134	4.2	1075	4	US-09-252-991A-18387	Sequence 18387, A
33	134	4.2	3256	4	US-09-919-172-98	Sequence 98, Appl
34	134	4.2	3256	4	US-09-976-594-22	Sequence 22, Appl
35	132.5	4.1	491	1	US-07-903-103-2	Sequence 2, Appl
36	132.5	4.1	491	1	US-08-044-619A-2	Sequence 2, Appl
37	132.5	4.1	491	1	US-08-283-911-2	Sequence 3, Appl
38	132.5	4.1	491	1	US-08-245-500A-3	Sequence 3, Appl
39	132.5	4.1	491	1	US-08-390-546-3	Sequence 3, Appl
40	132.5	4.1	491	1	US-08-390-479A-3	Sequence 3, Appl
41	132.5	4.1	491	1	US-08-557-393-3	Sequence 3, Appl
42	132.5	4.1	491	1	US-08-390-516C-3	Sequence 3, Appl
43	132.5	4.1	491	1	US-08-390-517A-3	Sequence 3, Appl
44	132.5	4.1	491	1	US-08-390-515A-3	Sequence 3, Appl
45	132.5	4.1	491	2	US-08-801-718-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-07-853-913-2
Sequence 2, Application US/07853913
Patent No. 5338839
GENERAL INFORMATION:
APPLICANT: McKay, Ronald D.G.
TITLE OF INVENTION: Nestin Expression As An Indicator of
TUMOR OR INFECTION: Neuroepithelial Tumors
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/853,913
FILING DATE: 19920319
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/660,412
FILING DATE: 22-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,803
FILING DATE: 25-OCT-1990
APPLICATION NUMBER: US 07/201,762
FILING DATE: 02-JUN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/180,548
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4641AA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1805 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULAR TYPE: protein

US-07-853-913-2

Query Match 6.1%; Score 197.5; DB 1; Length 1805;
 Best Local Similarity 21.2%; Pred. No. 8.3e-08;
 Matches 143; Conservative 99; Mismatches 223; Indels 211; Gaps 35;

QY 2 DDKG--HLSNEAPKAIPTSKKFRKTKWGFRTTIAKRGAGDTADPSQQPOQHLS 58
 DB 486 EDEGQIWEIWEKADIEVVENSSAQT-----QESGLDTBYTDSQQPLQKR-T 534
 QY 59 LARSROPRTREVEREFLTVARRGKNVPVSLSDS-----SEPTSTVTVDVETASGSYE 114
 DB 535 LVALBEPPLMSLTKQVETRA-----GKENCNSTBGLHGLTBGPBKQIPLSLKKNV 590
 QY 115 SSSSEIRSG-PV--SDSISKEHPASSER---AKG-----GEEEDTSDSDSDG----- 155
 DB 591 SKRTLENGVPVLSLLGKEDRTTDEQELMSPKGTLLKRFPSLSGSGEVAVRPSKEGMLBSM 650
 QY 156 -----LTLKELQN-----RLRRKRRQEP----- 173
 DB 651 TAFKESQHPLEGPAGABDQMLRLLVKEQDSPPRSPREBDOBAQCPLOKNEBPLQYBBA 710
 QY 174 ---VERSL-RGSQNRRLRKGREDSAGTSVQIGSABDRPLCKQEPSPASQSPVSCSET 228
 DB 711 EGQILRLIKESQESLRSP--EEEDQENG-----RSLQK-----NQEPLGYBBA 754
 QY 229 DD--IENOLEGKATOG--NTEENPREAGRP---KPECEVYDPNALYCTICQPHNNRPM 279
 DB 755 EDQMLERLIKESQESLKGPEEMQR-IGKPLERENKSLRYLEENQBTVPLESRRQRL 813
 QY 280 ICCDRCER-----WTHQDQVGS-----BARGLU 303
 DB 814 RSLVEEBEQRIVKPLKVSQDSLSGLAEENVPRLYLEBDQICNLSLBDTKSLGSL 873
 QY 304 LBRNGEDVYCPCNTLIQV-----ODETNGSATNEDSGCRRSVAGDGTD 346
 DB 874 EDRNGDSIIIPQBSERQVSLRPPBEDQRIYHILKESQEPERSSEBEGWAKRSLEGEN 933
 QY 347 CTSIGTVEKSGEDQIKRIEKNANPSGK---KLLKIPQVVEAPGAPKICIGPCSSV 402
 DB 934 HESLSLSEVER---EDQVSESQLEKESQDSGKSLDESDQETGRL-----EKENALSLRSL 984
 QY 403 ACPDYSVCNDCLIKLAATMRPLSSGKQ-KTKPEKXKTKPEKSLPKCSVQVQIKIS 461
 DB 985 ACPDQ--EBQKLEQETQOTLR--AVGNEQMAVSPPEKV--DPR--PLG--NDQ 1030
 QY 462 SVHKRLASEKREN-----PVKKV-----MLASRSETSGKAACGSSTPSWASDH 505
 DB 1031 EIASRSLGKNGESVLSLKEKGIETVKSLETRIIETLETAEDLERKKSIDTQEPMLSTRV 1090
 QY 506 NYNAVPEKPEKPTAL 521
 DB 1091 ARETVPEDEPPGSL 1106

RESULT 2
 US-09-141-047-8
 ; Sequence 8, Application US/09141047A
 ; Patent No. 6043085
 ; GENERAL INFORMATION:
 ; APPLICANT: Walker, David H.
 ; APPLICANT: Yu, Xue-Die
 ; TITLE OF INVENTION: Ehrlichia canis 120-kDa Immunodominant Antigenic
 ; TITLE OF INVENTION: Protein Gene
 ; FILE REFERENCE: D6143
 ; CURRENT APPLICATION NUMBER: US/09/141,047A
 ; CURRENT FILING DATE: 1998-08-27
 ; NUMBER OF SEQ ID NOS: 15
 ; SEQ ID NO 8
 ; LENGTH: 688
 ; TYPE: PRT
 ; ORGANISM: Ehrlichia canis
 ; FEATURE:

OTHER INFORMATION: Amino acid sequence of 120 kDa immunoreactive
 ; OTHER INFORMATION: protein.
 ; US-09-141-047-8

Query Match 5.1%; Score 166; DB 3; Length 688;
 Best Local Similarity 21.4%; Pred. No. 1.1e-05;
 Matches 128; Conservative 84; Mismatches 211; Indels 176; Gaps 30;

QY 8 SNEBAPKAIPTSKKFRKTKWGFRTTIAKRGAGDTADPSQQPOQHLSLRSGROPK 67
 DB 90 SSSSEVKKVSKTSK-----ESTPEVKAR-----DLQPAVDGSIHSSS--EVEKVS 135
 QY 68 KTERVEEFLTVARRGKNVPVSLSDSSEPTSTV-----TVETAS 109
 DB 136 KTSK--ESTPEVK--ABDLPVAVDSVHSSSEVGEKVSSTSKENTPEVKAEDLPVAV 191
 QY 110 EGSVE--SSSEI-----RSGP-----VSDSLGKHPAS--SEKAKGEBER 146
 DB 192 DSGIHSSESVGEKVSSTSKENTPEVKAEDLPVAVDSV--EHSSEVGEKVSSTSKER 249
 QY 147 DTSDDSDGLTLKELQNLRLRRRQEPVRSLSGQNLKTKRREDSAPTSVQIGSAR 206
 DB 250 NTPVAVABDLP-----QPAVDGSIHSSSEVGEK--VSTSKESTPEV-KAR 293
 QY 207 QDRPLCKQRPBASQSPVSG--SETTDIENQLEK-----ATQANTEENPREAG----- 252
 DB 294 DLQPAVDSEHSSSEVGEKVSSTSKENTPEVKAEDLPVAVDSVSEHSSSEVGEKVSST 353
 QY 253 ---KPECEVYDPNALYCTICQPHNNRPMICCDRCBWFHQDQVGISEARGLLBRNGR 309
 DB 354 SKESTPEVKAEDLP-----LQ-----AVDSIR--HS-----SSEVGEKVSSTSK 391
 QY 310 DYICPCNTLIQVDETFNGSATNEDSGCRRSVAGDGTCTSIGVBOKSGD--QGIKRI 367
 DB 392 ESTPEVKAEDLPVAVDSVSEHSSS-----EVEKVSSTSKENTPEVKAEDLPVAVDSV 447
 QY 368 EKAANPSGK-----KLLKIPQVVEAPGAPKICIGPCSSVAPDPSVCSNDC 414
 DB 448 EHSSEVGEKVSSTSKENTPEVKAEDLPVAV-----DG 481
 QY 415 ILKHAATM--RFLSGKQKTKPEKVKTKPE--KESLPKCSVQVQIKISSVK-----R 466
 DB 482 SVHSSSEVGEKVSSTSKENTPEVKAEDLPVAVDSVSEHSSSEVGEKVSSTSKESTPE 541
 QY 467 LASEKREPYKVMALASRSTSGK--EAACSSSTPSMASPHNMYNAYKPEKPTALSP 523
 DB 542 VKAEDLPVAVDSVSEHSSSEVGEKVSSTSKESTPEVKA-----VQPAVDGMPVPLNP 595

RESULT 3
 US-08-769-309A-5
 ; Sequence 5, Application US/08769309A
 ; Patent No. 5741890
 ; GENERAL INFORMATION:
 ; APPLICANT: Scott, John D.,
 ; APPLICANT: Nauert, Brian J.,
 ; APPLICANT: Klauk, Theresa M.
 ; TITLE OF INVENTION: Protein Binding Domains of Gravin
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gershtein, Murray & Borun
 ; STREET: 6300 Sears Tower/233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/769,309A

FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 574189and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/33451
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1780 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULAR TYPE: protein
US-08-769-309A-5

Query Match 5.1%; Score 165.5; DB 1; Length 1780;
Best Local Similarity 20.4%; Pred. No. 4.8e-05;
Matches 133; Conservative 81; Mismatches 201; Indels 237; Gaps 32;

QY 3 DKGHLSN---BEAPKA-----IKPTSK-----FRKTM---GPRRT----- 32
DB 136 DDGQENNNIBOIPSSSMLBELTQPTESQANDIGKFKVGFVPTVKQKTERPDTV 195
QY 33 ---TIKREGAGDTA---DPS-----EQQPOHNLILRSGRQPKTERVEEFL 76
DB 196 QLLTVKDEBGAAGAGHODPSLGAAGAAKSESEKQ-----STKKEBETL 242
QY 77 TVVRARGKNVPSLSDSEPTSTVTDVETASGSSVSSSEIRSGPVSDSLCK----- 130
DB 243 KRGSHALISPPA---ESGQAVBECKEGBEKEKPSKASPSPTSPTSSTGTFKPF 299
QY 131 -----EHPASSERAKGGEEDTSDSDGILTKELQNLRLRRKROEPVERSLGSON 183
DB 300 TQWAGMRKKTSPFRPK--EDEVEASE-----KKKQEP----- 331
QY 184 RLKRRKREED-SAETSGVOIGSABODRPLCKQEP--BASQGPVSQSEETDI-----ENQLBG 237
DB 332 --EKVDTEBDGALVASEKLTASBQNH--QEPASAHPRLSATYKVELPSEBQVSG 386
QY 238 KATQGTENBPREAGPKP--ECBVYDPMALYCIQRPNNRRPMICCDRCBEMFHGDCVGI 296
DB 387 --SQGPSEB---KPAFLATEVFDEKI-----EVHQBKVVA-----EVHV 420
QY 297 SEARGHLERNGEDYICNPCTILQVODETNGSANTNODSGCSVGADGTDCTSIGTVEOK 356
DB 421 STVEBTEBQKTE-----VEETAGSVPAEELVGMADAPQEAEPAKELVTKET 468
QY 357 --SGED--OGI-----KGRIRKANPSGKKLKI-----FQVVEAPGAPKCI 395
DB 469 CVSGEDPTQADLSPEDEKLSKRPBGVSVSEVEMLSQGEKMKVQGSPLKLTFTSTGLKLS 528
QY 396 G-----PGSSVAOPDSVYCSNDCILHGA 421
DB 529 GKQKQKCGGEBEGEHTQVPADSPDQBPQKSSASSPPEPEITC-----LEKGLA 583
QY 422 TWRLTSSGKQTKPKPEKVK-----TKPEKPSLPKCSVQVQGIKISSVHKRLASEGRNP 475
DB 584 EVQDQGEABEGATSDQEKKBGVTPWASFKKMTVPKRV-----RRPSBSDKEDB 633
QY 476 VKKVULASRSETSGKEACBSSTPWSADHNNANVPEKPEKPTALSPITLS 527
DB 634 LDKVKSATLSST-----BSTASEKQEMKGSVEBPPEPKKRVDTSVS 677

RESULT 4
US-08-994-570-5
Sequence 5, Application US/08994570
Patent No. 6090929
GENERAL INFORMATION:
APPLICANT: Scott, John D.,

APPLICANT: Nauert, Brian U.,
APPLICANT: Klauck, Theresa M.
TITLE OF INVENTION: Protein Binding Domains of Gravin
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum
STREET: 6300 Sears Tower/233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/994,570

FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090929and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/33451
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1780 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULAR TYPE: protein
US-08-994-570-5

Query Match 5.1%; Score 165.5; DB 3; Length 1780;
Best Local Similarity 20.4%; Pred. No. 4.8e-05;
Matches 133; Conservative 81; Mismatches 201; Indels 237; Gaps 32;

QY 3 DKGHLSN---BEAPKA-----IKPTSK-----FRKTM---GPRRT----- 32
DB 136 DDGQENNNIBOIPSSSMLBELTQPTESQANDIGKFKVGFVPTVKQKTERPDTV 195
QY 33 ---TIKREGAGDTA---DPS-----EQQPOHNLILRSGRQPKTERVEEFL 76
DB 196 QLLTVKDEBGAAGAGHODPSLGAAGAAKSESEKQ-----STKKEBETL 242
QY 77 TVVRARGKNVPSLSDSEPTSTVTDVETASGSSVSSSEIRSGPVSDSLCK----- 130
DB 243 KRGSHALISPPA---ESGQAVBECKEGBEKEKPSKASPSPTSPTSSTGTFKPF 299
QY 131 -----EHPASSERAKGGEEDTSDSDGILTKELQNLRLRRKROEPVERSLGSON 183
DB 300 TQWAGMRKKTSPFRPK--EDEVEASE-----KKKQEP----- 331
QY 184 RLKRRKREED-SAETSGVOIGSABODRPLCKQEP--BASQGPVSQSEETDI-----ENQLBG 237
DB 332 --EKVDTEBDGALVASEKLTASBQNH--QEPASAHPRLSATYKVELPSEBQVSG 386
QY 238 KATQGTENBPREAGPKP--ECBVYDPMALYCIQRPNNRRPMICCDRCBEMFHGDCVGI 296
DB 387 --SQGPSEB---KPAFLATEVFDEKI-----EVHQBKVVA-----EVHV 420
QY 297 SEARGHLERNGEDYICNPCTILQVODETNGSANTNODSGCSVGADGTDCTSIGTVEOK 356
DB 421 STVEBTEBQKTE-----VEETAGSVPAEELVGMADAPQEAEPAKELVTKET 468
QY 357 --SGED--OGI-----KGRIRKANPSGKKLKI-----FQVVEAPGAPKCI 395
DB 469 CVSGEDPTQADLSPEDEKLSKRPBGVSVSEVEMLSQGEKMKVQGSPLKLTFTSTGLKLS 528
QY 396 G-----PGSSVAOPDSVYCSNDCILHGA 421

Db 529 GKKQKRGCGGBRSGEHTQVPADSPDSBQKGBSSASSPPEPBITC-----LEKGLA 583
 Qy 422 TWRPLSSGKQKTKPEKVK-----TKPEKPSLPKCSYQVVGIKISSVHKRLASEKREP 475
 Db 584 EYQOQDABEABGATSDCKRGVTPWASPCKQVTPKRV-----RRPEESDKEDB 633
 Qy 476 VCKQMLASSBTSBGKAAACSSSPSMAASHNRYAVPEPEKPTALSPILLS 527
 Db 634 LDKVKSATLSST-----ESTASSEMCKGVSVEPPEKPKRVTSVS 677

RESULT 5

US-09-257-179-80
 ; Sequence 80, Application US/09257179
 ; Patent No. 6410709
 ; GENERAL INFORMATION:
 ; APPLICANT: Ruben et al.
 ; TITLE OF INVENTION: 29 Human Secreted Proteins
 ; FILE REFERENCE: P2015P1
 ; CURRENT APPLICATION NUMBER: US/09/257,179
 ; EARLIER FILING DATE: 1999-02-25
 ; EARLIER APPLICATION NUMBER: PCT/US98/17709
 ; EARLIER FILING DATE: 1998-08-27
 ; EARLIER APPLICATION NUMBER: 60/056,270
 ; EARLIER FILING DATE: 1997-08-29
 ; EARLIER APPLICATION NUMBER: 60/056,271
 ; EARLIER FILING DATE: 1997-08-29
 ; EARLIER APPLICATION NUMBER: 60/056,247
 ; EARLIER FILING DATE: 1997-08-29
 ; EARLIER APPLICATION NUMBER: 60/056,073
 ; EARLIER FILING DATE: 1997-08-29
 ; NUMBER OF SEQ ID NOS: 128
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 80
 ; LENGTH: 238
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-257-179-80

Query Match 5.1%; Score 164; DB 4; Length 238;
 Best Local Similarity 36.2%; Pred. No. 3.4e-06;
 Matches 29; Conservative 16; Mismatches 11; Indels 24; Gaps 3;

Qy 238 KATQGTENPREAGKPECEVYDPAALYCTROPNN-NRFMICDRCGEWTHGCVGI 296
 Db 51 RAQGGSSER-----LYCICRTPYDSQFYIGCDRCQNTWTHGRCVGI 91
 Qy 297 SEARGRLERKEDYICPNC 316
 Db 92 LOSEBELT-----DEVVCPQC 107

RESULT 6

US-08-296-791-6
 ; Sequence 6, Application US/08296791
 ; Patent No. 624337
 ; GENERAL INFORMATION:
 ; APPLICANT: St. Geme III, Joseph W.
 ; TITLE OF INVENTION: Haemophilus Adherence and Penetration
 ; TITLE OF INVENTION: Protein
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSES: Plehr, Hohbach, Test, Albritton & Herbert
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/296,791
 ; FILING DATE: 25-AUG-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Trecartin, Richard F.
 ; REGISTRATION NUMBER: 31,801
 ; REFERENCE/DOCKET NUMBER: A-59941/RPT/RMS
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989
 ; TELEFAX: (415) 398-3249
 ; TELEX: 910 277299
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1848 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; US-08-296-791-6

Query Match 4.8%; Score 153.5; DB 3; Length 1848;
 Best Local Similarity 21.7%; Pred. No. 0.00055;
 Matches 124; Conservative 71; Mismatches 208; Indels 169; Gaps 25;

Qy 8 SNEEAKAIKPTSKERKTMGPRRTIARREGAGTDEAPSE-----OQPOCHNTSLRR 61
 Db 1072 STETAPKSDATVTE-----NPNSESVPSTETTKVAENPPENETVAK 1114
 Qy 62 SGRO-----PKTERVEEPLTV-----RRGKNVVPVLSDBSSEPTSTVTV 106
 Db 1115 NQGEATEPTPNQNEVAKEDQPTVEAVTQNEATQSGKQLEPTQAFKSEPTTS-VTVSE 1173
 Qy 107 TASEGVSSESSERTSPVSDSLCKEHPASERKAGGEKEHETSDSDGLTKELQRLR 166
 Db 1174 MQPEKTVSOSTE-----DKVVE--KREKXK--VETEFQKAVQ--VTSKBPQKQAR 1219
 Qy 167 RREKQRPVRSLSGQNRRLKRRBEDSAFTQSGAARQ--DRPLCKQEP----- 217
 Db 1220 PAPERVPTTNABEALQOTQPTTVAALBTISPNKSPABETQPSBKTAABVTPVVSZ 1279
 Qy 218 --ASQGPVOSQETDIDENQLGKATQNTENPRE--AGKP-----KPECEVYDPAALYC 268
 Db 1280 NTAQO-PTETETRAKYKXKKTQEVPPQVASEPKQBPAPAKQVTKQBPABENVL-- 1336
 Qy 269 ICRQPHNNRPMICDRCCEWTHGDCVIGSEARGLLERNGEDYICNCTILQVQDETNGS 328
 Db 1337 -----TKNVGEPO-----POAQPOQST 1355
 Qy 329 A---TNEODSGCKSVGADGTC-----TSIGTVBQSGEDQGIKGRIBKANPSGKK 377
 Db 1356 AVPTTETANSPKAPAPQAPQAPQTEPARENVSTNTEBQSO-----TSATYSTBQ 1408
 Qy 378 KLRIFQVVEAPAPAPPCIGGCSVAQPSVYCSNDCLIKHAATNRFLSSGKQKTPK 437
 Db 1409 PAKETSSNTEQAPREHSIMTG-----SALTW--TETAKSDKQ 1445
 Qy 438 EKYKT--KPEKPSLPKCSYQVVGIKIS--SVHGLASEKRENPKYKWLARSSTSGK 430
 Db 1446 METVTENDRQPEANVTADNSVANNSSSSKSPRRRSVQPKETSAEBTTVASTOBTTVD 1505
 Qy 491 EACGESSTPMASDHVNAVKP---EKPEKPT 519
 Db 1506 NSV---STPEPNSKRTRRSVQTNSTYEPVLPPT 1534

RESULT 7

US-09-839-996-6
 ; Sequence 6, Application US/09839996
 ; Patent No. 6642371
 ; GENERAL INFORMATION:
 ; APPLICANT: St. Geme III, Joseph W.
 ; Falkow, Stanley

Db 1506 NSV---STPKPRSRTRRSVQTNSTYEPVELPT 1534

RESULT 9

PCT-US95-10661A-6

Sequence 6, Application PC/TUS9510661A
GENERAL INFORMATION:
APPLICANT: Washington University, et al.
TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSER: Flehr, Hobbach, Test, Albrighton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10661A
FILING DATE: 16-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Treccartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: FP-59941/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1848 amino acids
TYPE: amino acid
TOPOLOGY: unknown
PCT-US95-10661A-6

Query Match 4.8%; Score 153.5; DB 5; Length 1848;
Best Local Similarity 21.7%; Pred. No. 0.00055;
Matches 124; Conservative 71; Mismatches 208; Indels 169; Gaps 25;

QY 8 SNEAPKAIKPTSKKFRKTWGRRTTIARBGAGDTADPSE-----QQPOQHLSLRK 61
DB 1072 STETAPKSDTATQTE-----NPNSESVPSEETREKVAENPQENETVAK 1114
QY 62 SGRQ-----PKRTEVBEFLTV-----RRRGKNVVSLEDSSEPTSTTVDR 106
DB 1115 NEQEAETPTPQNGEVAKEQDPVEANQTQNEATQSEKGETQNTATTSSEPTES-VTVSE 1173
QY 107 TASEGSVSSSIRSGPVSDSLGKHEPASSEKAKGGBREEDTSDSDGLTKELQNRK 166
DB 1174 NQEKTVQSTE-----DKVYVE---KEEKAK--VETRETOYKAPQ--VTSKEPRKQAR 1219
QY 167 RKREGEPEVRSIRGQNRKRRKREDSAEFGSVQIGSAEQ--DRPLCKQEPF----- 217
DB 1220 PAPEEVPTDYNAEQAALQOTPTTVAALETSPNSKPAEETQQSEKTNABPVTPVASE 1279
QY 218 --ASQGPVSQSEITDIDENLEGKATQNTENPRE---AGKP-----KPECEVYDENALYC 268
DB 1280 NTATQ-PTRETTATVKEKETOBOVAVASQSPKQOPAKQAOTKQAQAFARENVL-- 1336
QY 269 ICRQPHNNRFMIICDRCEBEPHGDGCVGISSEARGLLERNAGDYICPNCITLLQVODETNGS 328
DB 1337 -----TTKNVGEHQ-----PQAQPTQST 1355

QY 329 A---TNEQDSGCRSVGADGTD-----TSIGTVBQSGEDQGIKRIERKANPSGKK 377
DB 1356 AVPTTGETNANSKPAKPAKQPAQKPTQEPARENVSTVNTBPSQ-----TSATYSTQ 1408
QY 378 KLIKFPVEAPGAPKICIGPCCSSVAQPDVSVCNDILKHAATNRFISGREGQTKPK 437
DB 1409 PAKETSSNVEQPAPEPNSINTG-----SATIM---TETAEKSDKQ 1445
QY 438 EKVKT---KPEKSLPKCSVOVGIKIS---SVHKLASKEKENPVKYMALASRSTSGK 490
DB 1446 METVTENDRQPAANTVADNSVANNSESSSKRRRRSVSQPKETSALETTVAJSTOETVAD 1505
QY 491 EAACSSSTPSWASDENYNAVKP---EKPEKPT 519
DB 1506 NSV---STPKPRSRTRRSVQTNSTYEPVELPT 1534

RESULT 10

US-08-285-440-5

Sequence 5, Application US/08285440
Patent No. 5532337

GENERAL INFORMATION:
APPLICANT: Ken'ichiro HAYASHI et al.
TITLE OF INVENTION: POLYPEPTIDE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSER: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Displaywrite
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/285,440
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/858,947
FILING DATE: March 27, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:

IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-285-440-5

Query Match 4.7%; Score 151.5; DB 1; Length 532;

Best Local Similarity 20.9%; Pred. No. 0.00013; Matches 120; Conservative 66; Mismatches 233; Indels 155; Gaps 23;

30 RRTTAKRGAGDTBADPESQPOOHNLSTRSGRQPK-RTREVEEFLTVRRGKKNP 88
10 RRSIALSQIAYQRNDDEBEARER---RRARQERLRQOEESLGQVTDQVAVNAQ 65
89 VSLBDSSEPTSTVTDVETASBGSVSSSEIRSGPVSDSLGKEHPASSSEKAKGSEEDT 148
66 NSVPDEAKTTTNTNOVE-----GDDEAFLRLARRRERQ- 102
149 SDSDSGLTLKLNRLRKREQEPVERSLRGSQNRRLKREEDSAFSGVQISABED 208
103 -----KRLQALERQKEFDP---TTDASLSLPRRQNDTAEN---ETTERE 144
209 RPLCKQEPASQSPVSOSET-----DIEVOLGKATOGTENTENPREAG----- 252
145 E---KSEBROERYIBETETVTKSYQKNDWRDAENKEDKEKEBEBEKKPKRSGIGENQ 201
253 -----KPKPE-----CEVYDNNALYICICROPH-NNRPMICCDRCSEWPH 290
202 IKDEKIKDKKPKKEHVKSFMORRKGFTTEVKSQNGEFMTKLTHTENTFSRPGGRAS--- 257
291 GDCVGISERAG-----RLERNGE-----DYICPNCITLIQVODETN 326
258 ---VTKRABGAPQVEAGRLLELRRRGETSEBEPKIKQKQOQALALEBELKCKRER 314
327 GSATNBODSGCSVAGDGTCTSIGTVEQKSGEDGIGRIKRAANPSGKKLKIFQPVY 386
315 RKLVEEQRKQKQEBAD-----RKLREBEBKRLKEIIRRRABAAEKQKQKMEDEL 366
387 BAPGAP-KCIGPGCSSVAQPDVSVCSNDCLK-----HAATYKFLSSGKEQTKPKPE 438
367 SDDKPFKCFPKGSLKIEBRAFLENKSVQKSGVKSSTHQAIAIVSKIDSRLEQYTSAL 426
439 KVK-----TKPEKPSLPKCSVQVIGIKIISVHKRLASBKKNPVKXWMLAS--RSETSGKEA 492
427 GTSKAKPKPKPAASDLP-----VPAEGVRIKSNMBKGNVPSPTAAGTNPATGLKV 479
493 ACBSSTPSMAS---DHNTAAVPEKPEKPTALSP 523
480 GVSSRIINEMUTPTDGN-----KSPAP-KPSDLRP 508

RESULT 11
US-08-630-349-5

Sequence 5, Application US/08630349
Patent No. 5739008
GENERAL INFORMATION:
APPLICANT: Ken'ichiro HAYASHI et al.
TITLE OF INVENTION: POLYPEPTIDE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSER: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 144 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,349
FILING DATE: April 10, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/285,440
FILING DATE: August 4, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/858,947
FILING DATE: March 27, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Warren W. Cheek, Jr.
REGISTRATION NUMBER: 33,367
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-630-349-5

Query Match 4.7%; Score 151.5; DB 1; Length 532;

Best Local Similarity 20.9%; Pred. No. 0.00013; Matches 120; Conservative 66; Mismatches 233; Indels 155; Gaps 23;

30 RRTTAKRGAGDTBADPESQPOOHNLSTRSGRQPK-RTREVEEFLTVRRGKKNP 88
10 RRSIALSQIAYQRNDDEBEARER---RRARQERLRQOEESLGQVTDQVAVNAQ 65
89 VSLBDSSEPTSTVTDVETASBGSVSSSEIRSGPVSDSLGKEHPASSSEKAKGSEEDT 148
66 NSVPDEAKTTTNTNOVE-----GDDEAFLRLARRRERQ- 102
149 SDSDSGLTLKLNRLRKREQEPVERSLRGSQNRRLKREEDSAFSGVQISABED 208
103 -----KRLQALERQKEFDP---TTDASLSLPRRQNDTAEN---ETTERE 144
209 RPLCKQEPASQSPVSOSET-----DIEVOLGKATOGTENTENPREAG----- 252
145 E---KSEBROERYIBETETVTKSYQKNDWRDAENKEDKEKEBEBEKKPKRSGIGENQ 201
253 -----KPKPE-----CEVYDNNALYICICROPH-NNRPMICCDRCSEWPH 290
202 IKDEKIKDKKPKKEHVKSFMORRKGFTTEVKSQNGEFMTKLTHTENTFSRPGGRAS--- 257
291 GDCVGISERAG-----RLERNGE-----DYICPNCITLIQVODETN 326
258 ---VTKRABGAPQVEAGRLLELRRRGETSEBEPKIKQKQOQALALEBELKCKRER 314
327 GSATNBODSGCSVAGDGTCTSIGTVEQKSGEDGIGRIKRAANPSGKKLKIFQPVY 386
315 RKLVEEQRKQKQEBAD-----RKLREBEBKRLKEIIRRRABAAEKQKQKMEDEL 366
387 BAPGAP-KCIGPGCSSVAQPDVSVCSNDCLK-----HAATYKFLSSGKEQTKPKPE 438

```

Db      367 SDDKPFKCFPKGSSILKIEERAFINKSGVSKSTHQAALSKIDSLRLEYTSAIE 426
Qy      439 KVK-----TKPEKSLPKCSVOVGITCSVHKRLABKREKNPKYKMLAS--RSESGHEA 492
Db      427 GTKAKAPKPAASDLF-----VPAEGVRNIKSMWKEGVNFSPTAAGTPEKETAALKV 479
Qy      493 ACESSTPSMAS---DHNYNAVPEKPEKPTALSP 523
Db      480 GVSSRINEMLTKTPDGN-----KSPAP-KPSDLRF 508

```

RESULT 12

```

US-09-976-594-726
; Sequence 726, Application US/0976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 726
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 4255277CD1
US-09-976-594-726

```

```

Query Match      4.6%; Score 149.5; DB 4; Length 2468;
Best Local Similarity 20.5%; Pred. No. 0.0018;
Matches 134; Conservative 87; Mismatches 250; Indels 183; Gaps 31;

```

```

Qy      7 LSNBAPALKFT-----SKEFRKTWGRRTT---IAREGAGDTBADPSEQPOOHNLIS 58
Db      861 IEDBEKLKETEPEVAVVLOKEREVTKGPASPEDEGTTTBGBCBCHQPEBLEP----- 914
Qy      59 LRRSGRQKRTERVAEPLTTVARRKQANPVLSLDSSEPTSTVTVDETASGRVSSSE 118
Db      915 VERQG-----VDIEKP-----EDGAGFESSSE-----TGDIYBKAE 947
Qy      119 IRSGVSDSLGKEHPASSEKAKGGEEDTSDSDGLTLKELQNLKREKRE----- 170
Db      948 TEAEPEDDEBEHYCVASAKHSPTEDEBSAKADAY-----IRKRSVASGDOR 999
Qy      171 -QEPVRSGL-RGSQNLKRRKREDSATG-----SVQIGSAE 206
Db      1000 AEDMDRAIKKEARQSEBEADBEKADARBEYEPEKMEADYVMVAVDAAREGAE 1059
Qy      207 QD-----RPLCQ-----PEAS-----QGVSSQSETDIDENQ-----LEKATQGN 243
Db      1060 EGYGLTTPYKQLAQSFGREPASSIHDETLFGSGSESEATASDEENREPQPEFTATISGY 1119
Qy      244 TEENPRBAGKPECEVVDNALYCICRQPHNN-----RFMICCDRCSEWFKGD----- 292
Db      1120 TOSTIBISSEPTPMDEMSTPRD---VMSDETNNBETESPQGFVNITKYESSLYQGEYK 1176
Qy      293 -----CVGISARGLLERNGEDYICPNCITL---QVDETNGSATNEODSGCRSYGADG 344
Db      1177 PADVTPUNGFSRG-SEKTATDQKDYNAASATISPPSSMEBDFSRSALEADAYCSEYKAST 1235
Qy      345 T-----DCSGTGVBEKSGSDGIGKRIEKAANPSGKKLKIFQPVYAEAPGAKCIGPGCS 400
Db      1236 TLIDIOSISAVSSSEKVSPS---KSPSLSPPSPPLEKTPLEGRSVNLSLTNEILVNSAE 1291
Qy      401 SVAQPSDVSVCNDCLIKHAAA-----TWRLS-----SGKEQKTK--PRE 438

```

```

Db      1232 AEVAVSEVAVQVEVEHCAPEDEKTLLEVSPSQVTCAGHTPYQSPTEKSSHLPT 1351
Qy      439 KVKTEPEKSLPKCSVOVGIRKISV--HKRLAS-----EKRENPVKVNLASRS-ET 487
Db      1352 VIKRP-----AVPFSFESDADKDNERRASVSPMBPVPDSDSPLEKVLSPLRSPPL 1403
Qy      488 SGKEACSES--STPSMASDHNYNAVPEKPEK--PTALSP--LISKTTHPK 534
Db      1404 IGSBSAYSFSLADDKASGAGSEPFEEKSGKQSPDQVPSVSEMTSTSLYODK 1457

```

RESULT 13

```

US-08-893-852A-3
; Sequence 3, Application US/08893852A
; Patent No. 6080558
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,852A
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0341 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 657 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 53041
US-08-893-852A-3

```

```

Query Match      4.6%; Score 148; DB 3; Length 657;
Best Local Similarity 19.4%; Pred. No. 0.00036;
Matches 128; Conservative 64; Mismatches 241; Indels 226; Gaps 26;

```

```

Qy      41 GPTBADPSEQPO-QHNLSLRSGRQPKRT---BRVERP-----LTTVRRGRKX 85
Db      75 GHTSESGSPESQAARLCLVAESSPPTWGLSNVDENAPGQDDLRKEMERTAGTA 134
Qy      86 NV-PVSLDESEPTSTVTVDETASGRVSSSEIRSGFVSDS----- 127
Db      135 TLQPAQLQADRELAVARBEQVAEPAY-PTSQLEGGAEKREBDEGTATKYQASNASTA 193
Qy      128 -----LGR-EHPASSSEKAKGGEEDTSDSDGLTLKELQNLKRRKREBQ 172

```



```

Db      194 PGYKSTVPFLGAEHQATBE--KGTENKADPSNSPSSGSHSRAMEVYSREKPKQEGEA 251
      173 PVERSLRGSQNLRLKRRREDSAEFGSVQIG-----SABOD 208
      252 KVEARAGCGHCRRNABABGGPRTTFVCTGNAFLKAWYRRPGEDTEEDNSDSABED 311
      209 RPLCKQEPBAS-----OGPVQSESTDJLENOLBGATQNTBENPREAKGPBECE 259
      312 TAQGTATHTSAFLKAWYRRPGEDTEEDNSDSABEDTAQTGAT-----356
      260 VYDPNALYICQHPNNRMICCDRCBWFHGDVCGISBARGRLERNGEDYICPNCTIL 319
      357 -----PHTSAFL-----KAWY-----RPGEDTEEDNSDLD 382
      320 QVQDETNGSATNEODSGCRS--VGADGTDCTSIGTVEOKSGEDOGIKGRIKXANPSGKK 377
      383 SAEEDTAQTGATPHTSAFLKAWYRRPGED-----TEEDNSDLSABEDTAQTGATPHTSP 437
      378 KLK--IPOP-----VVEAPGAPKCIGPCSSVAOPDSVYCSNDCLIKHAATMRFLSSG 429
      438 FLKAWYRRPGEDTEEDTEEDSENVAPGDSSTASOSGCL-----479
      430 KEQKTKPEKVK-----TKPEK--PSLPKCSV-----454
      480 QPQRCLEPKETKGRGBBPLFOVAFTLPGEKDESPWAAPKFLRLQRLRLPKAPTRDQD 539
      455 -QVGKISSVH-----KRLASERKENPVKKV-----MLASRSETSGKE 491
      540 PEIPLKARVHFAEKTYVFLAVMAGPAAARGPEQPARDRSRARIRLAQAEBKLGKY 599
      492 AACSESTPSMASDHYNAVKEPKPEKPTALSPILLSKCTYHP-KAGFPGPSHHLGCLG 549
      600 LTPDSRARAMARLRNPSLPQSE-PRSSSEATPLTDVTPPLPSETPPSPSLYLGRRG 657

RESULT 14
US-08-821-818-3
; Sequence 3, Application US/08821818
; Patent No. 616877
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: IDENTIFICATION OF THE PROGRESSION
; TITLE OF INVENTION: ELLEVATED GENE-3 AND USES THEREOF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,818
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Chan, Albert Wei-Kit
; REGISTRATION NUMBER: 36,479
; REFERENCE/DOCKET NUMBER: 51523
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 657 amino acids
; TYPE: amino acid
; STRANDBESS: single
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: protein
; US-08-821-818-3
Query Match      4.6%; Score 148; DB 3; Length 657;
Beet Local Similarity 19.4%; Pred. No. 0.00036;
Matches 128; Conservative 64; Mismatches 241; Indels 226; Gaps 26;

      41 GDTADPSEBQPPQ-QHNLSLRSGROPKRT-----ERVEEP-----LTTVRRGKK 85
      75 GETESGSPBQQAQRCLVAESSPEFTWGLSNVDENAKPGODLREKMERATAGKA 134
      86 NV-PVSLDSEPTSTYTDVETASRGVSSSEIRGVSDS-----127
      135 TLQPAQLAGADRRLGAEVVARBEGVAEPAY-PTSOLEGPAENBDEFTVQYQASAA51A 193
      128 -----IGK--RHPASSERKAGGEEREDSDSGTLIKXONLRRRBOE---112
      194 PGYKSTVPFLGAEHQATBE--KGTENKADPSNSPSSGSHSRAMEVYSREKPKQEGEA 251
      173 PVERSLRGSQNLRLKRRREDSAEFGSVQIG-----SABOD 208
      252 KVEARAGCGHCRRNABABGGPRTTFVCTGNAFLKAWYRRPGEDTEEDNSDSABED 311
      209 RPLCKQEPBAS-----OGPVQSESTDJLENOLBGATQNTBENPREAKGPBECE 259
      312 TAQGTATPHTSAFLKAWYRRPGEDTEEDNSDSABEDTAQTGAT-----356
      260 VYDPNALYICQHPNNRMICCDRCBWFHGDVCGISBARGRLERNGEDYICPNCTIL 319
      357 -----PHTSAFL-----KAWY-----RPGEDTEEDNSDLD 382
      320 QVQDETNGSATNEODSGCRS--VGADGTDCTSIGTVEOKSGEDOGIKGRIKXANPSGKK 377
      383 SAEEDTAQTGATPHTSAFLKAWYRRPGED-----TEEDNSDLSABEDTAQTGATPHTSP 437
      378 KLK--IPOP-----VVEAPGAPKCIGPCSSVAOPDSVYCSNDCLIKHAATMRFLSSG 429
      438 FLKAWYRRPGEDTEEDTEEDSENVAPGDSSTASOSGCL-----479
      430 KEQKTKPEKVK-----TKPEK--PSLPKCSV-----454
      480 QPQRCLEPKETKGRGBBPLFOVAFTLPGEKDESPWAAPKFLRLQRLRLPKAPTRDQD 539
      455 -QVGKISSVH-----KRLASERKENPVKKV-----MLASRSETSGKE 491
      540 PEIPLKARVHFAEKTYVFLAVMAGPAAARGPEQPARDRSRARIRLAQAEBKLGKY 599
      492 AACSESTPSMASDHYNAVKEPKPEKPTALSPILLSKCTYHP-KAGFPGPSHHLGCLG 549
      600 LTPDSRARAMARLRNPSLPQSE-PRSSSEATPLTDVTPPLPSETPPSPSLYLGRRG 657

RESULT 15
US-09-052-753B-3
; Sequence 3, Application US/09052753B
; Patent No. 6472520
; GENERAL INFORMATION:
; APPLICANT: Paul B. Fisher
; TITLE OF INVENTION: Progression Elevated Gene-3 and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: A34608-B
; CURRENT APPLICATION NUMBER: US/09/052,753B
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: PCT/US98/05793
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 08/812,818
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 657
; TYPE: PRT
; ORGANISM: rat

```


PS Claim 8; Fig 1D; 27pp; English.

XX The invention provides nucleic acids encoding the human and murine death
CC inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be
CC expressed by standard recombinant methodology. The DIO-1 polypeptides,
CC agonists and antagonists are used as a medicament for treating diseases
CC characterized by an alteration in cell death or by hyperproliferation,
CC e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
CC tumours, malignant tumours or hyperproliferative skin disorders. They are
CC also useful in the treatment of metabolic, proliferative or inflammatory
CC conditions. The present sequence represents the murine DIO-1 polypeptide
XX
SQ Sequence 614 AA;

Query Match 100.0%; Score 3228; DB 3; Length 614;
Best Local Similarity 100.0%; Pred. No. 1,4e-215;
Matches 614; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MODKGHLISNEBAPKAIKPTSKERFKTWGFRRTTIARBGAGDTADPSEQOPQOHNLSIR 60
DB 1 MODKGHLISNEBAPKAIKPTSKERFKTWGFRRTTIARBGAGDTADPSEQOPQOHNLSIR 60
QY 61 RSGRQPKRTREVEEFLITVRRRGKKNVPVSLDSEPTSTVTVDVETASGVSSESSIR 120
DB 61 RSGRQPKRTREVEEFLITVRRRGKKNVPVSLDSEPTSTVTVDVETASGVSSESSIR 120
QY 121 SGPVSDSLCKEHPASSEKAKGSEEBTSDSDGLITKELQNRRLRKRRQEPVERSLNG 180
DB 121 SGPVSDSLCKEHPASSEKAKGSEEBTSDSDGLITKELQNRRLRKRRQEPVERSLNG 180
QY 121 SGPVSDSLCKEHPASSEKAKGSEEBTSDSDGLITKELQNRRLRKRRQEPVERSLNG 180
DB 121 SGPVSDSLCKEHPASSEKAKGSEEBTSDSDGLITKELQNRRLRKRRQEPVERSLNG 180
QY 181 SGNRLRKRRREDSATFSVOIGSARQDRPLCKQEPBASQSPVSGSETDIEQLGKAT 240
DB 181 SGNRLRKRRREDSATFSVOIGSARQDRPLCKQEPBASQSPVSGSETDIEQLGKAT 240
QY 241 QGNTENPREAKPKPECEVYPNALYICICROPNNRPMICDRCCEWTFHGDVCGISBAR 300
DB 241 QGNTENPREAKPKPECEVYPNALYICICROPNNRPMICDRCCEWTFHGDVCGISBAR 300
QY 301 GLLBERNGEDYICPNCTIIQVODETNGSATNBODSGCRSVGADTCTSIGTWEQSGSD 360
DB 301 GLLBERNGEDYICPNCTIIQVODETNGSATNBODSGCRSVGADTCTSIGTWEQSGSD 360
QY 361 GQIKRIRKRNAPSGKKIKITROPVVEAGAPKICPGSSVAPQSPVCSNDCLIKHAA 420
DB 361 GQIKRIRKRNAPSGKKIKITROPVVEAGAPKICPGSSVAPQSPVCSNDCLIKHAA 420
QY 421 ATRRFLSSGKEQKTKPEKTKPEKFSLPKCSVOVGIIKISVHKRLASKEKNPVKRW 480
DB 421 ATRRFLSSGKEQKTKPEKTKPEKFSLPKCSVOVGIIKISVHKRLASKEKNPVKRW 480
QY 481 LASRSETSGKEAKCSSTPSWASDHNYNAVKPEKPEKPTALSTTLISKCTYHKAGPPGP 540
DB 481 LASRSETSGKEAKCSSTPSWASDHNYNAVKPEKPEKPTALSTTLISKCTYHKAGPPGP 540
QY 541 SHHGGCGCLISTRTVGVVLIYASSSLPARSRYODASGQVPLPSLMSLSGFTLSGCVG 600
DB 541 SHHGGCGCLISTRTVGVVLIYASSSLPARSRYODASGQVPLPSLMSLSGFTLSGCVG 600
QY 601 LMLBAISYSPFRPW 614
DB 601 LMLBAISYSPFRPW 614

```

RESULT 2
AAB93638
ID AAB93638 standard; protein; 562 AA.

XX AAB93638;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:13130.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX BP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI, 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-

XX length cDNAs defined in the specification, and for the detection and/or

XX diagnosis of the abnormality of the proteins encoded by the full-length

XX cDNAs.

XX Claim 8; SEQ ID NO 13130; 2537pp + Sequence listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-

XX length cDNAs defined in the specification. Where a primer set comprises:

XX (a) an oligo-dt primer and an oligonucleotide complementary to the

XX complementary strand of a polynucleotide which comprises one of the 5602

XX nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX oligonucleotide comprises at least 15 nucleotides and the combination of

XX the 5'-end sequence/3'-end sequence is selected from those defined in the

XX specification. The primer sets can be used in antisense therapy and in

XX gene therapy. The primers are useful for synthesizing polynucleotides,

XX particularly full-length cDNAs. The primers are also useful for the

XX detection and/or diagnosis of the abnormality of the proteins encoded by

XX the full-length cDNAs. The primers allow obtaining of the full-length

XX cDNAs easily without any specialized methods. AAH03166 to AAH13628 and

XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893

XX represent human amino acid sequences; and AAH13629 to AAH13632 represent

XX oligonucleotides, all of which are used in the exemplification of the

SQ Sequence 562 AA;

Query Match 65.4%; Score 2110; DB 4; Length 562;

Best Local Similarity 75.0%; Pred. No. 5.2e-138;

Matches 420; Conservative 32; Mismatches 94; Indels 14; Gaps 7;

```

QY 1 MODKGHLISNEBAPKAIKPTSKERFKTWGFRRTTIARBGAGDTADPSEQOP--QOHNLS 58
DB 1 MODKGHLISNEBAPKAIKPTSKERFKTWGFRRTTIARBGAGDTADPSEQOPQOHNLSIR 60
QY 59 LRRSGQPKRTREVEEFLITVRRRGKKNVPVSLDSEPTSTVTVDVETASGVSSESSIR 118
DB 61 LRRSGQPKRTREVEEFLITVRRRGKKNVPVSLDSEPTSTVTVDVETASGVSSESSIR 120
QY 119 IRSGPVSDSLG-KEHPASSEKAKGSEEBTSDSDGLITKELQNRRLRKRRQEPVERSL 177
DB 121 TRSGPOSASTAVAKERASSEKAKGSDHDHDTSDSDGLITKELQNRRLRKRRQEPVER 180
QY 178 LRGSGNRLRKRRREDSAT--GSVOIGSARQDRPLCKQEPBASQSPVSGSETDIEQLG 236
DB 178 LRGSGNRLRKRRREDSAT--GSVOIGSARQDRPLCKQEPBASQSPVSGSETDIEQLG 236

```

```

Db      181 LKGIQSLRKRRREBPAETVGSBASDTVBGLP-SKQEPENDGVVSQAQKDRRESKLE 239
Qy      237 GKATQGTENPRNAPKPKPCERYDPPNALYCICQPPNNRMICCDRCBEMPHGDCVGI 296
Db      240 GKAAQDIDKEBPDLGRPECEBGYPNALYCICQPPNNRMICCDRCBEMPHGDCVGI 299
Qy      297 SEARGRLLENGEDYICPNCITLLQVODETNGSATNBODSGRSVAGDGTDCISIGTVBOK 356
Db      300 SEARGRLLENGEDYICPNCITLLQVODETHSETADQQAARRPGDADDTCTISIGTIEOK 359
Qy      357 SGEBOGIGRIEKANPSSGKKKTKIPOPVBARPAKPCIGPCSSVAQPDSDVYCSNDCIL 416
Db      360 SSEDGIGKRIEKANPSSGKKKTKIPOPVBARPAKPCIGPCSSVAQPDSDVYCSNDCIL 419
Qy      417 KHAATMRFLSSGKEOKTKPKKVKTKPKESLPCQSVQVGIKISSVHKRLASERREMPV 476
Db      420 KHAATMRFLSSGKEOKTKPKKVKTKPKESLPCQSVQVGIKISSVHKRLASERREMPV 479
Qy      477 KK-VMLASRSETSGKEAACBSSTPSMASDHNYNVAKPEKPEPTALSPTLISKCTYHPKA 535
Db      480 KKAVVVPARSRBALGKEAACBSSTPSMASDHNYNVAKPEKPTAP---SPSLAYKCMYHLGV 536
Qy      536 GPPGPSHHL-----GGCLGL 550
Db      537 GLDPSRSFWAIAPWACPGL 556

RESULT 3
AAV67579
ID      AAV67579 standard; protein; 562 AA.
XX
AC      AAV67579;
XX
DT      19-JUN-2000 (first entry)
XX
XX      Human death inducer-obliterator 1 (DIO-1) polypeptide.
XX
KW      Death inducer-obliterator 1; DIO-1; cell death; cancer; tumour; human;
KW      autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
KW      antineumatic; antiinflammatory; antiproliferative.
XX
OS      Homo sapiens.
XX
FH      Key Location/Qualifiers
FH      Region 165..172
FH      /note="NLS sequence"
FH      Region 185..193
FH      /note="NLS sequence"
FH      Region 271..288
FH      /note="zinc finger motif"
FH      Region 293..320
FH      /note="zinc finger motif"
XX
PN      WO200015787-A1.
XX
PD      23-MAR-2000.
XX
PF      10-SEP-1999; 99WO-GB003019.
XX
PR      10-SEP-1998; 98SR-00003069.
XX      17-SEP-1998; 98US-0100873P.
XX
PA      (CMSJ) CONSELJO SUPERIOR INVESTIGACIONES CIENTIF.
PA      (BANN/) BANNERMAN D G.
XX
PI      Alonso CM, Domingo DG, Grandien A, Leonardo E, Martinez P;
XX
DR      WPI: 2000-271426/23.
XX      N-PSSB; AA290578.
XX
PT      New DNA encoding human and murine death inducer-obliterator 1
PT      polypeptides, useful in the treatment of cancer, autoimmune diseases,
PT      diabetes, rheumatoid arthritis, benign tumors, malignant tumors and

```

```

PT      hyperproliferative skin disorders.
XX
PS      Claim 6, Fig 1C; 27pp; English.
XX
CC      The invention provides nucleic acids encoding the human and murine death
CC      inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be
CC      expressed by standard recombinant methodology. The DIO-1 polypeptides,
CC      agonists and antagonists are used as a medicament for treating diseases
CC      characterized by an alteration in cell death or by hyperproliferation,
CC      e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
CC      tumors, malignant tumors or hyperproliferative skin disorders. They are
CC      also useful in the treatment of metabolic, proliferative or inflammatory
CC      conditions. The present sequence represents the human DIO-1 polypeptide
XX
SQ      Sequence 562 AA:
XX
Query Match      65.3%; Score 2108; DB 3; Length 562;
Beat Local Similarity 75.0%; Fred. No. 7.1e-136;
Matches 420; Conservative 32; Mismatches 94; Indels 14; Gaps 7;
Qy      1 MDDKGLSNEBAPKAIKPTSKPFRKTMGFRRTTIARSGAGPTEADPSBOCP--OOHNLIS 58
Db      1 MDDKGDPSNEBAPKAIKPTSKPFRKTMGFRRTTIARSGAGAEADPLAPPPOOGLGIS 60
Qy      59 LRRSGROPKTERVEEPLTVARRRGNVPSVLSDSSPTSTVTVDVTASGVSSESSB 118
Db      61 LRRSGROPKTERVEEPLTVARRRGRSMVSLSDSGPTSCPATDAETASGVSASABE 120
Qy      119 TRSGPVSLSLC-KEHPASSERAKGEEBBDTSDSDGLTKELONRLARKRQERVERS 177
Db      121 TRSGPOSASTAVKEKPPASSEKVKGGDDHDSDSDGLTKELONRLARKRQERTERP 180
Qy      178 LRGSQNLRLKRRREDSABT-GSVQIGSABQDRPLCKQSPBASQSPVSGSFTDLENQLE 236
Db      181 LKGIQSLRKRRREBPAETVGSBASDTVBGLP-SKQEPENDGVVSQAQKDRRESKLE 239
Qy      237 GKATQGTENPRNAPKPKPCERYDPPNALYCICQPPNNRMICCDRCBEMPHGDCVGI 296
Db      240 GKAAQDIDKEBPDLGRPECEBGYPNALYCICQPPNNRMICCDRCBEMPHGDCVGI 299
Qy      297 SEARGRLLENGEDYICPNCITLLQVODETNGSATNBODSGRSVAGDGTDCISIGTVBOK 356
Db      300 SEARGRLLENGEDYICPNCITLLQVODETHSETADQQAARRPGDADDTCTISIGTIEOK 359
Qy      357 SGEBOGIGRIEKANPSSGKKKTKIPOPVBARPAKPCIGPCSSVAQPDSDVYCSNDCIL 416
Db      360 SSEDGIGKRIEKANPSSGKKKTKIPOPVBARPAKPCIGPCSSVAQPDSDVYCSNDCIL 419
Qy      417 KHAATMRFLSSGKEOKTKPKKVKTKPKESLPCQSVQVGIKISSVHKRLASERREMPV 476
Db      420 KHAATMRFLSSGKEOKTKPKKVKTKPKESLPCQSVQVGIKISSVHKRLASERREMPV 479
Qy      477 KK-VMLASRSETSGKEAACBSSTPSMASDHNYNVAKPEKPEPTALSPTLISKCTYHPKA 535
Db      480 KKAVVVPARSRBALGKEAACBSSTPSMASDHNYNVAKPEKPTAP---SPSLAYKCMYHLGV 536
Qy      536 GPPGPSHHL-----GGCLGL 550
Db      537 GLDPSRSFWAIAPWACPGL 556

RESULT 4
AAV60219
ID      AAV60219 standard; protein; 1191 AA.
XX
AC      AAV60219;
XX
DT      06-NOV-2001 (first entry)
XX
DR      Human protein SBQ ID NO 3865.
XX
PT      Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW      vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

```

KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KM nervous system disorder; arthritis; inflammation.
 XX Homo sapiens.
 XX MO200157190-A2.
 PD 09-AUG-2001.
 XX 05-FEB-2001; 2001MO-US004098.
 PF 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00663561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX (HYSB-) HYSEQ INC.
 PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y,
 PI Ma Y, Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZM;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX WPI; 2001-476283/51.
 DR N-PSDB; AAK53352.
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
 in diagnosis and gene therapy.
 XX Claim 20; Page 455-456; 6221pp; English.
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX
 XX Sequence 1191 AA:
 SQ
 Query Match 55.5%; Score 1792.5; DB 4; Length 1191;
 Best Local Similarity 70.6%; Pred. No. 1.5e-115;
 Matches 358; Conservative 32; Mismatches 74; Indels 43; Gaps 6;

QY 361 OGKGRIEKAANPSGKKKIKLP----- 384
 DB 321 OGKGRIEKAANPSGKKKIKLP----- 380
 QY 385 VVEAPGAPKCTIGPGCSVAQPDVYCSNDCTIKHAAATRTPLSGKEQTKREKXTKP 444
 DB 381 VTEAPGASKCTIGPGCSVAQPDVYCSNDCTIKHAAATRTPLSGKEQTKREKXTKP 440
 QY 445 EKPSPKCSVOYGITSSVHRRLASEKENPVKK-VMLASRETSQKAACSSSTPSMAS 503
 DB 441 EKPSPKCSVOYGITSSVHRRLASEKENPVKK-VMLASRETSQKAACSSSTPSMAS 500
 QY 504 DHNYNAVPEKPEKPTALSPTLISKCT 530
 DB 501 DHNYNAVPEKPTALSPTLISKCT 524
 RESULT 5
 ABG22389
 ID ABG22389 standard; protein; 775 AA.
 XX
 AC ABG22389;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #22380.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001MO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSB-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS86576.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 XX Claim 20; SEQ ID NO 52748; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probe, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic

CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 775 AA;

Query Match	54.4%	Score 1756.5	DB 4	Length 775
Best Local Similarity	57.1%	Pred. No. 2.7e-113		
Matches 378	Conservative 45	Mismatches 126	Indels 113	Gaps 14

QY	1	MDDGHLSNBEAPKAIKPTLSKEFKWTGGRRTIARKEAGDTEAPSEQOP--QOHNLS	58
Db	131	MDDKHDPNBEAPKAIKPTLSKEFKWTGGRRTIARKEAGDTEAPSEQOP--QOHNLS	150
QY	59	LRRSGRQPKTERVEEPLTVRRKQNPVSLBEDSSEFTSTVTDVETASBGSVSSSE	118
Db	191	LRRSGRQPKTERVEEPLTVRRKQNPVSLBEDSSEFTSTVTDVETASBGSVSSSE	250
QY	119	IRSGVPSVSLG--KEHPASSBKAKGHEEDTDSDSGLTLKEIQRNLRRKQEPERS	177
Db	251	TRSGQPSASTAVKEHPASSBKAKGHEEDTDSDSGLTLKEIQRNLRRKQEPERS	310
QY	178	LRRGQNRRLRKQREEDSAET--GSVQISABODRPLCKQEPKASQGVPSOSETDIENOLB	236
Db	311	LKGISRLRKQRRREGPALETGVSBSADTVBGLP--SKQEPENDQGVASQAGKODRESKLB	369
QY	237	GKATQNTAEENPREKGRPECEVYDPNALYICICQPHNNRPMTCCDRCEMHFGCVGI	286
Db	370	GKAQODIDBERBGDGRPPBCEGYDPNALYICICQPHNNRPMTCCDRCEMHFGCVGI	429
QY	297	SEARGRLERNGEDYICPNCTILQVODETNGSATNODSGCSVVG--ADGTDCSITGVEQ	355
Db	430	SEARGRLERNGEDYICPNCTILQVODETHSEFADQBAKMGDLGDADGDFPSIGNIRS	489
QY	356	K--SGEDGIGIKRIKRAANPSGKKLTKIQP-----VYEAPCA	391
Db	490	RLSLRRPRLGLKRIKRAANPSGKKELKIQPVLEIAVSRISIAFTLLHCTISCVIARPCA	549
QY	392	PKCIGPGSSVAPQPSVYCSNDICILKHAATYRPLSSGGEOKTKPEKVTYKPEKSLPK	451
Db	550	SKCIGPGCHVAPQPSVYCSNDICILKHAATYKPLSSGGEOKTKPEKMTMKREKESLPLK	609
QY	452	C-----SVQVG-----ITISVYHKLASEKRENPVAKVM	480
Db	610	CGAOPVQWILLISVQLGTSSGLQVTLCPPLTEHNKMLKFSVYHAGETAGPRARHRLKTL-	668
QY	481	LANSSETSGRKAACSSSTPVSASDHNYNNAVKKPEKPLALSPTLISKCTYHKKAGAPCP	540
Db	669	-----NKTYGACHCNPLRP-----LPCLT-----PCQPHRS	693
QY	541	SHHLLGCILGLSTRV-----LGVLVILVASSSLPARSRYODASGPQVYLPSELMSLSGV	593
Db	694	SPSLVIGAGSLPRHVAHNLBPCLAI-----CPLARRLSPARGPLASGDDLTWGYKVV	744
QY	594	PL 595	
Db	745	LL 746	

RESULT 6
AAM79235
ID AAM79235 standard; protein; 647 AA

XX
 AC AAM79235;
 XX
 XX 06-NOV-2001 (first entry)
 DT
 XX
 XX Human protein SEQ ID NO 1897.
 DE
 XX
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;
 KM

KW nervous system disorder; arthritis; inflammation.

05 **Homo sapiens.**

PN WO200157190-A2.

PD 09-AUG-2001
YY

03-FBB-2001; 2001MO-03004056-XX

PR 27-APR-2000; 2000US-00560875.

PR 19-JUL-2000; 2000US-00620325.

PR 15-SBP-2000; 2000US-00663561.

PR 30-NOV-2000; 2000US-00728422.

PA (HYSB-) HYSEQ INC.
XX

PI Ma Y., Zhao Q.A., Wang D., Wang
PI Tang H., Liu C., Drmannac R.I.

XX

DR N-PSDB; AAK52368.

PT Nucleic acids encod

XX
PC

XX
CC The invention relates

cytokine, cell pro

polynucleotides are

e.g., stem cell gro

CC activin/inhibin ac

The invention relates to polypeptides (AAK51455-AAK53435) and the encoded polypeptides (AAW763123-AAW80302) that exhibit activity eliciting cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polypeptides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activity/inhibit activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the sequence listing were missing at the time of publication

SQ Sequence 647 AA;

Query Match	52.5%	Score 1693.5;	DB 4	Length 647;
Best Local Similarity	70.4%	Pred. No. 5.1e-109;		
Matches 340;	Conservative 28;	Mismatches 72;	Indels 43;	Gaps 6

87VP6LSBDBSBPTSTWTDVETASBBSVSSSEIRSRSPVSDLG-KHBPASSKANGKBBR 155

146 EDPDSDDGLTAKELONRLRRKKBQEPYRSLRSONRLAKRKRBDSDAER-GSVQLGS 204

Q7 205 AEDRPLKQBPBASQSPVQSSTDDIINQLEGKATQNTNENPFRAGKPKRECEVYDNP 26

265 ALYLCICRPHNNRPMICDRCBEMFHGDCGISLARGLLBRNGBDYICPNCTIIQYDE 324

```

240 THISBTADQBAKHPGDADTDCISIGTLEOKSSDEOGIKGRIKRAPSGKXKLQFP 299

```

```

Oy 365 -----VVEAPGAPKCIIGPCSSVAOPDSV 408
Db 300 GPGVPYPTQPLVMQVLEIAVRSISAFITLHICISCVIEAPGASKCIIGPCCHVAOPDSV 359
Oy 409 YCSNDCILKHAATATRPFLSSGKOKTKPEKVKTKPEKSLPKCSQVQVGIKISVHKRLA 468
Db 360 YCSNDCILKHAATATKPLSSGKOKTKPEKVKTKPEKSLPKCSQVQVGIKISVHKRPA 419
Oy 469 SEKRNPNYK-VMLASRSETSKKAACSSSTPSMASDHNHNAVKPEKPEKPTALSPITLS 527
Db 420 PEKKTITTKAVVAPRSALGKEAACESTPSMASDHNHNAVKPEKTAAP---SPSILLY 476
Oy 528 KCT 530
Db 477 KST 479

RESULT 7
ABG22388
ID ABG22388 standard; protein; 330 AA.
XX
AC ABG22388;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #22379.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PI 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-MAR-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
DR N-PSDB; AAS86575.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostic, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 52747; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostic, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this

```

```

CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 330 AA;
XX
XX Query Match 37.9%; Score 1224.5; DB 4; Length 330;
XX Best Local Similarity 70.7%; Pred. No. 7.9e-77;
XX Matches 232; Conservative 21; Mismatches 36; Indels 37; Gaps 2;
XX
Oy 232 ENQLEKATQGTENPNPBAKPKPCBCEYVDNALYCTIRQPHNNRPMICCRCEWTFHG 291
Db 1 BSKLEKAAQDIKDEBEPDIAKPKPCBCEYVDNALYCTIRQPHNNRPMICCRCEWTFHG 60
Oy 292 DCVGISSEARGRLERNGRDYICPNCTIIQVODBTNGSATNEODSGCRSVGADGDTCTSIG 351
Db 61 DCVGISSEARGRLERNGRDYICPNCTIIQVODBTNGSATNEODSGCRSVGADGDTCTSIG 120
Oy 352 TVEOKSGEDQIGRIEKAAPSGKKQLKIFQF----- 384
Db 121 TIEQSSSDQIGRIEKAAPSGKKQLKIFQFPGVPYPTQPLVMQVLEIAVRSISAF 180
Oy 385 -----VVEAPGAPKCIIGPCSSVAOPDSVYCSNDCILKHAATATRPFLSSGKOKTK 435
Db 181 TILHICISCVIEAPGASKCIIGPCCHVAOPDSVYCSNDCILKHAATATKPLSSGKOKTK 240
Oy 436 PEKVKTKPEKPSLPKCSQVQVGIKISVHKRLASEKRNPNYK-VMLASRSETSKKAAC 494
Db 241 PEKKNMPEKPSLPKCSQVQVGIKISVHKRPAPEKKTITTKAVVAPRSALGKEAAC 300
Oy 495 BSSTPSMASDHNHNAVKPEKPEKPTALS 522
Db 301 BSSTPSMASDHNHNAVKPEKTAAPSPVT 328

RESULT 8
AAB43724
ID AAB43724 standard; protein; 181 AA.
XX
AC AAB43724;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated protein sequence SEQ ID NO:1169.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW antidiabetic; antiaesthetic; antineumatic; antiaesthetic; antiviral;
KW antiinflammatory; antihypertoid; antiallergic; antibacterial; cardiac;
KW dermatological; neuroprotective; thrombolytic; coagulant; nocitropic;
KW vasotropic; antiporiatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening.
XX
XX Homo sapiens.
XX
XX WO200055350-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000MO-US005882.
XX
XX 12-MAR-1999; 99US-0124270P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen Ca, Ruben SM;
XX
XX WPI; 2000-587533/55.
XX
XX N-PSDB; AAC77933.
XX

```


PT Novel isolated nucleic acids comprising sequences encoding peptides
 PT useful for treating or diagnosing e.g. cancer.
 XX
 XX Claim 11, Page 1790-1791; 2352pp; English.
 CC AACT78607 to AACT78448 encode the human cancer associated proteins given in
 CC AAB43338 to AAB44239. The proteins can have activities based on the
 CC tissues and cells the genes are expressed in. Example of activities
 CC include: cytostatic; proliferative; vulnerable; immunomodulator;
 CC anti-diabetic; anti-neurotic; anti-neurotic; anti-neurotic;
 CC anti-inflammatory; antitumor; antiallergic; antibacterial; antiviral;
 CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
 CC nocotropic; vasotropic; antipsoriatic and antiangiogenic. The
 CC polynucleotides and polypeptides can be used for preventing, treating or
 CC ameliorating medical conditions and diagnosing pathological conditions.
 CC Polynucleotides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AACT78449 to
 CC AACT8457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention
 XX
 XX Sequence 181 AA:

Query Match 22.9%; Score 740; DB 3; Length 181;
 Best Local Similarity 85.1%; Pred. No. 1.6e-43;
 Matches 131; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

252 GKPRCECVYDNNALYCTIGROPHNNRPMCCRCCEMFHGDGCTSEARGLLEKRGEDY 311
 11 GKPRCECVYDNNALYCTIGROPHNNRPMCCRCCEMFHGDGCTSEARGLLEKRGEDY 70
 312 ICPNCTIIQVODETNGSATNBODSGCRSVGADGCTCTSIGTVEOKSGEDQIKRIEKAA 371
 71 ICPNCTIIQVODETNGSATNBODSGCRSVGADGCTCTSIGTVEOKSGEDQIKRIEKAA 130
 372 NPSGKKKLIKFPVVEAPGAPKCTIGPGCSVAOP 405
 131 NPSGKKKLIKFPVVEAPGAPKCTIGPGCSVAOP 164

RESULT 9
 AAG22387
 ID AAG22387 standard; protein; 167 AA.
 XX
 AC AAG22387;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #22378.
 XX
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PP 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 XX
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSB-) HYSBQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;

XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAG86574.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostic, forensic, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 52746; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probe, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 167 AA:

Query Match 22.7%; Score 732; DB 4; Length 167;
 Best Local Similarity 83.1%; Pred. No. 5.2e-43;
 Matches 138; Conservative 9; Mismatches 19; Indels 0; Gaps 0;

301 GRLLRNGEDYICPNCTIIQVODETNGSATNBODSGCRSVGADGCTCTSIGTVEOKSGED 360
 1 GRLLRNGEDYICPNCTIIQVODETNGSATNBODSGCRSVGADGCTCTSIGTVEOKSGED 60
 361 OGKRIEKAAPNSGKKKLIKFPVVEAPGAPKCTIGPGCSVAOPPSVCSNDCLIKHAA 420
 61 OGKRIEKAAPNSGKKKLIKFPVVEAPGAPKCTIGPGCSVAOPPSVCSNDCLIKHAA 120
 421 ATMRPLSSGKEQYKPKKVKTKPKKPSLPKCSVQVGTIKSSVHKR 466
 121 ATMRPLSSGKEQYKPKKVKTKPKKPSLPKCSVQVGTIKSSVHKR 166

RESULT 10
 ABP05496
 ID ABP05496 standard; protein; 108 AA.
 XX
 AC ABP05496;
 XX
 DT 24-JUN-2002 (first entry)
 XX
 DE Human ORFX protein sequence SEQ ID NO:10974.
 XX
 KW Human: open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myaesthesia gravis.
 XX
 OS Homo sapiens.
 XX

PW WO200192523-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 29-MAY-2001; 2001WO-US010836.
 XX
 PR 30-MAY-2000; 2000US-0206132P.
 XX 29-AUG-2000; 2000US-0228716P.
 XX
 PA (CUPRA-) CURAGEN CORP.
 XX
 PI Shinkens RA, Leach MD;
 XX
 DR WPI: 2002-106308/14.
 DR N-PSDB; ABR21248.
 XX
 PT Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders.
 XX
 PS Disclosure; SEQ ID NO 10974; 1037pp; English.
 XX
 CC The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABR15762 to ABR27252 encode the human ORFX
 CC proteins given in ABR00010 to ABR11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, disorders related to organ
 CC transplantation, neurodegenerative disorders, diabetes mellitus, systemic
 CC lupus erythematosus, cardiovascular diseases, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage. N.B. The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 108 AA;
 Query Match 17.3%; Score 559; DB 5; Length 108;
 Best Local Similarity 99.1%; Pred. No. 3.1e-31;
 Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 357 SGEQGIKIRIKANPSPGSKGLKIFQPVVEAPGAKCIGPGSSVAQDPDVSNDCLL 416
 DB 1 SGEDEGIKIRIKANPSPGSKGLKIFQPVVEAPGAKCIGPGSSVAQDPDVSNDCLL 60
 QY 417 KHAATMRLSSGKQKTPKPKVKTKPKPSLPKCSVGVGKISSVH 464
 DB 61 KHAATMRLSSGKQKTPKPKVKTKPKPSLPKCSVGVGKISSVH 108
 RESULT 11
 ABB71487
 ID ABB71487 standard; protein; 2016 AA.
 AC ABB71487;
 XX 26-MAR-2002 (first entry)
 DT
 XX Drosophila melanogaster polypeptide SEQ ID NO 41253.
 DE
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KM

KM pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US0092231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 XX 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR N-PSDB; ABL15590.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 PS Disclosure; SEQ ID NO 41253; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABR57737-
 CC ABR72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2016 AA;
 Query Match 10.0%; Score 323.5; DB 4; Length 2016;
 Best Local Similarity 22.7%; Pred. No. 2.7e-13;
 Matches 160; Conservative 86; Mismatches 253; Indels 205; Gaps 30;
 QY 8 SNEEAPKAIK--PFSKEFRKTGFRRTI---AKREGAD-TEADPSEQQPQ----- 53
 DB 639 ADEEDIPKLVKMLESTRADIDPPVSLPTIIFAPBEKQIGNAAAAPNNIMDTANDEIT 698
 QY 54 ----OHNLSTARRSGRPKRTVERBEFLTVRRRGGKNVPSLDSSEPTSTVTVETAS 109
 DB 699 ADFLGHVGLIEDNQ-FAEAVVKQVLAFT-ERGLDALVSMPTSLIEVDVPOAHTNLTP 756
 QY 110 EGSV-BSSEETISGVP-----SDSLCKEHPASSERKXGGEEDTDSDSGLTKLEON 163
 DB 757 NASLTERPACSMSTSLPACSTPSPRSVAASTPTSAKVRYGRV-----TYLPTI-- 805
 QY 164 RARRRREQSPVRSRSGNQRLKKRREDSAVTSGVQISARQD---RPLCKQEPBAS 219
 DB 806 -----EAPTTAKRAQPPSAPGMAATSSSDAGULSGESSLDASINQPL----- 850
 QY 220 QGFSVQSSTDDIENQLEKATQNTENREKAPK-----PFC--- 258
 DB 851 -----MTSSLN--DSQPSGPKRPNRPSMARSTAPRRSKKLDSQNDPDASBS 901
 QY 259 --EYDPAALYICICROPHNRRPMICCDREBFWPHGCVGISARGLLRNEDDYCPNC 316
 DB 902 QEDDDDPNLTMCICROPHNRRPMICCDREBFWPHGCVGTAKMGDMENKGDWCKPKC 961
 QY 317 TILQVQDETNGSATNDSGCSVGADGTDCTSIGTVQKSGEGDQI--KGRIRKAPPS 374
 DB 962 VKRQ-----EENSOPRITTMVLTPTT--QPEQRPSTKVLTTTAEIVQVAAPS 1008
 QY 375 G-KKKLIFQPVVEAP-----GAPK-----CTPG----- 398

Db 1009 APRRTLPVVLTVASSPFRIRIPAKKPPYTGALSHQOQOOLNFTRLGSPSKRISSETLCLV 1068
Cc 399 -CSSVAGPDSVYGCNDILHMAATMPLSSGK-----EOKTKPERK 439
Cc 1069 VCRPAPSTSSVYCECEICRTKTAQSAIDHMAATKPLPQNAQASLLNNSPDAKKNKKDL 1128
Cc 440 VK-----TKPERK---PSLPKCSVOVGIKISSVHRLSEKRENPVKVMLASRSET 487
Cc 1129 FEDVLRADTVSKYERINVPFRKSGRVI TGHMAPSAH-QPFRKWIQENPSFEVLPSTGVQS 1167
Cc 488 SGKBAACESTPVSADHNTYNAV-----KPERKPEPALSPTLTSLKCTYHPKAGPQPSH 542
Cc 1188 ADABKRLIKGAPBAATSTSEPAVLGVAKKP---PEGPATLS-----HPQNTTVQASH 1236
Cc 543 HGGCGLSRTSRVGVVLVYASSSLPARSRVQDASGPQVFLPS 586
Cc 1237 QLG-----ISSVRPLAKKDKKKTPTTVOAPT 1262

RESULT 12
ID AAM39234 standard; protein; 1728 AA.
Cc AAM39234;
Cc 22-OCT-2001 (first entry)
Cc Human polypeptide SEQ ID NO 2379.
Cc Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
Cc peripheral nervous system; neuropathy; central nervous system; CNS;
Cc Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
Cc amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
Cc chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
Cc leukaemia.
Cc Homo sapiens.
Cc W0200153312-A1.
Cc 26-JUL-2001.
Cc 26-DEC-2000; 2000WO-US034263.
Cc 23-DEC-1999; 39US-00471275.
Cc 21-JAN-2000; 2000US-00488725.
Cc 25-APR-2000; 2000US-0052317.
Cc 20-JUN-2000; 2000US-00598042.
Cc 19-JUL-2000; 2000US-00620312.
Cc 03-AUG-2000; 2000US-00653450.
Cc 14-SEP-2000; 2000US-00662191.
Cc 19-OCT-2000; 2000US-00693036.
Cc 29-NOV-2000; 2000US-00727344.
Cc (HYSB-) HYSBQ INC.
Cc Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
Cc Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
Cc Zhou P, Goodrich R, Drmanac RT;
Cc WPI; 2001-442253/47.
Cc N-PSDB; AAI58190.
Cc Novel nucleic acids and polypeptides, useful for treating disorders such
Cc as central nervous system injuries.
Cc Example 4; SEQ ID NO 2379; 10078pp; English.
Cc The invention relates to human nucleic acids (AAI57798-AAI61369) and the
Cc encoded polypeptides (AAM38642-AA42213) with nootropic,
Cc immunosuppressant and cytostatic activity. The polynucleotides are useful
Cc in gene therapy. A composition containing a polypeptide or polynucleotide
Cc of the invention may be used to treat diseases of the peripheral nervous

Cc system, such as peripheral nervous injuries, peripheral neuropathy and
Cc localised neuropathies and central nervous system diseases, such as
Cc Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
Cc lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
Cc utilisation of the activities such as: immune system suppression,
Cc Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
Cc and thrombolytic activity, cancer diagnosis and therapy, drug screening,
Cc assays for receptor activity, arthritis and inflammation, leukaemia and
Cc C.N.S disorders. Note: The sequence data for this patent did not form
Cc part of the printed specification
Cc SQ Sequence 1728 AA;
Cc Query Match 7.4%; Score 239.5; DB 4; Length 1728;
Cc Best Local Similarity 20.5%; Pred. No. 1.5e-07;
Cc Matches 101; Conservative 75; Mismatches 163; Indels 153; Gaps 17;
Cc 90 SLEDSSEPTSTYDVTDFRAGSVSSSEIRIS--GPVDSLSGK-----BHPASSBKA 139
Cc 155 NLODDRNQSSSVSYLSEKSVSKTKPVISHKQMTTDAFKIVAAYEVTHSKTKNV 214
Cc 140 KGGEEEDTSDSDSGTLKELQN-----RLRRKR-KQSPV-----ERSLRSQNRRL 186
Cc 215 KSVKNTDVPBSQ-----QNFRRPVYKRAKKQIDKPKIQSCNSGVKYNQAHSTVL 265
Cc 187 KKRREDSAEFTGSVOI-----GSABQDRPLCKQBPB--ASQGPVSGSB----- 227
Cc 266 KKTLDQRL-----VQIFPDLTSLSDSKSHAHFGCLKBPHPAQYGHVSHSQKCHPQQ 321
Cc 228 -----TDIEMOLEGRATQGNTEENPREAGKPRPCE----- 259
Cc 322 QAPAKNTSHVGRLEHGVGEHFKEDKIKLKKPRKNUQPRQRSKSPSLDEPPLPIPD 381
Cc 260 -----VYDPNALVYCIGPQPHNRPMICCDRCCEWYFHGDVCSISA 299
Cc 382 NIATIRRGSDHSSSFESKTMPTPSKQCGFCKKPHGRNRPVGGCRDDPFGDCVLSLS 441
Cc 300 RGRLEHNGEDYICPNC-----TILQVQDETNGSATNBODSGCRSVAGADTCTSIGT 352
Cc 442 QAOQMGEBDEKVEYVCKCCABEDKTEILDPDLENGATVBPISGDKTM-----ECKERGL 496
Cc 353 VEOKSGE-----DQIKGRIRKANPSCGKKUKITQPVVEAAGAPKICPGGCSVAQPS 407
Cc 497 SKHTTNDRTKYIDTVYKHKVKILKRESGEGR-----NS 529
Cc 408 VYCSNDCLIKHAAATPRPLSSGKQKTEPKKVKTPKPKSLPKCSVOYGIKISVHKRL 467
Cc 530 SPCRNEIKKQWLAFLRKG-----QVLPFRSSSEKSKR--IPKSTTVYTCSEKASKRP 583
Cc 468 ASERKENPVKVV 479
Cc 584 THEKQEMKKKKV 595
Cc

RESULT 13
AAM79892
ID AAM79892 standard; protein; 1766 AA.
Cc AAM79892;
Cc 06-NOV-2001 (first entry)
Cc Human protein SEQ ID NO 3538.
Cc Human; cytokine; cell proliferation; cell differentiation; gene therapy;
Cc vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
Cc tissue growth factor; immunomodulatory; cancer; leukaemia;
Cc nervous system disorder; arthritis; inflammation.
Cc Homo sapiens.
Cc W0200157190-A2.
Cc

DB 841 SPCRDNEIKKQJLAPLRKG---OPVLRPRSSEKSEK--IPKESTVTCTGKASKPG 894
QY 468 ASEKRENVPYKV 479
DB 895 THKQEMKKKKV 906

Search completed: April 20, 2004, 21:18:22
Job time : 63 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 23, 2004, 07:20:53 ; Search time 7067.26 Seconds
(without alignments)
12114.300 Million cell updates/sec

Title: US-09-787-016a-2

Sequence: 1 tccgtgtgtagctcgtgaat.....cttgcctccttgcgtg99 2867

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estsba:*
2: em_esthm:*
3: em_esthm:*
4: em_esthm:*
5: em_esthm:*
6: em_esthm:*
7: em_esthm:*
8: em_esthm:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: em_estfun:*
15: em_estfun:*
16: em_estfun:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrt:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2828.2	98.6	4619	AK042474	Mus muscu
2	2775.8	96.8	4651	BC060249	Mus muscu
3	1482.4	51.7	1614	AK014422	Mus muscu
4	1398.8	48.8	1640	BC029110	Mus muscu

5	1010.8	35.3	2752	11	BC014615	BC014615 Homo sapi
6	835.6	29.1	3385	11	AK044919	AK044919 Mus muscu
7	780.8	27.2	874	13	B0522670	B0522670 AGENCOURT
8	702.2	24.5	796	12	B10782420	B10782420 B078113 602872420
9	701	24.5	682	13	BY713603	BY713603 BY713603
10	665.4	23.2	667	14	CF169555	CF169555 B0815C09-
11	621.4	21.7	691	10	BB637982	BB637982 BB637982
12	617.4	21.5	632	13	BQ771134	BQ771134 UI-M-PTO-
13	607.4	21.2	623	14	CP535718	CP535718 UI-M-GHO-
14	594.6	20.7	645	13	BB636983	BB636983 BB636983
15	591.6	20.6	746	13	BY752619	BY752619 BY752619
16	591.6	20.6	639	10	BB664208	BB664208 BB664208
17	587.4	20.5	613	14	CB518822	CB518822 UI-M-GHO-
18	579.4	20.2	693	10	BB292563	BB292563 BB292563
19	570.6	19.9	690	14	CA327345	CA327345 UI-M-PYO-
20	516.8	18.0	1086	9	AL578222	AL578222 AL578222
21	494	17.2	925	9	AL546121	AL546121 AL546121
22	482.8	16.8	486	28	CC156300	CC156300 XB629 Bay
23	481.4	16.8	492	12	BG342505	BG342505 602374063
24	469.4	16.4	491	12	B6342512	B6342512 602374072
25	467.2	16.3	670	12	B1653462	B1653462 603300427
26	450.4	15.7	521	14	CF167818	CF167818 B0790B04-
27	448.2	15.6	476	28	AZ390149	AZ390149 1M0151B22
28	438.4	15.3	891	13	BX350413	BX350413 BX350413
29	428	14.9	653	10	AM373892	AM373892 QV0-BT053
30	428	14.9	653	10	AM373892	AM373892 QV3-BT053
31	427.6	14.9	1093	9	AL555810	AL555810 AL555810
32	424.6	14.8	1018	9	AL550675	AL550675 AL550675
33	420.2	14.7	1033	9	AL580669	AL580669 AL580669
34	415.2	14.5	495	9	AL549786	AL549786 ve78e11.Y
35	412.6	14.4	506	14	CD541350	CD541350 B0231A09-
36	412.6	14.4	542	14	CD543468	CD543468 B0249C01-
37	402	14.0	760	10	BR298017	BR298017 601118492
38	398	13.9	773	12	BG545979	BG545979 602573278
39	392	13.7	1201	9	AL514138	AL514138 AL514138
40	389.8	13.6	850	14	CD364822	CD364822 UI-H-FT2-
41	389.2	13.6	772	10	BR299127	BR299127 601118879
42	386.8	13.5	650	10	BR298593	BR298593 601119753
43	386.4	13.5	868	10	BF306721	BF306721 601888796
44	385.8	13.5	1119	9	AL550431	AL550431 AL550431
45	385.2	13.4	498	12	BM247400	BM247400 K0801D11-

ALIGNMENTS

RESULT 1	AK042474	4619 bp	mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone: A63095A07 product: death inducer-oblierator-1, full insert sequence.
LOCUS	AK042474		
DEFINITION	AK042474		
ACCESSION	AK042474		
VERSION	AK042474.1	GI:26335118	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PubMed	10349636		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20493374		
PubMed	11042159		

REFERENCE
AUTHORS
3
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Komoto,H., Akiyama,Y., Nishi,K., Kitahara,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
Yamamoto,S., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujitake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,B., Watabiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN Integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
PUBMED
11076861

TITLE
4
The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

REFERENCE
AUTHORS
5
The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
AUTHORS
5 (bases 1 to 4619)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hirakawa,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanakawa,T.,
Kawai,H., Kawai,J., Kojima,Y., Kondo,S., Komuro,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akaiya,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Miyamatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
Trust/MRC building Addenbrookes Hospital Cambridge) whose
assistance we gratefully acknowledge.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://lancm.gsc.riken.go.jp/
Location/Qualifiers
1. 4619
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="CS:EBL/6J"
/db_xref="PANTOM_DB:A630095A07"
/db_xref="MGI:2407519"
/db_xref="taxon:10090"
/clone="A630095A07"
/tissue_type="Chymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="3 days neonate"
299. 2143
/note="unnamed protein product; death
inducer-obliterating-1 (MGI:MGI:1344352, GB|A0238332,
evidence: BLASTN, 99%, match=2599)

FEATURES
source

CDS

putative"
/codon_start=1
/protein_id="BAC31270.1"
/db_xref="GI:26335139"
/translation="MDDKGLHSNEAPAPAKPTSPKFKTWGFRRTTIAKRGADTE
VDPSEQPOOHNLRLSRGQPKRTYERETLYRRGKNVPSLSDSPSTSTV
TDVPSAGSVSSSRLSRGSDLSGKREDSALTSVQSGAEODPCLQKPEA
LQNLRLRRRGRBPVRLSRGSDNLRLSRGSDLSGKREDSALTSVQSGAEODPCLQKPEA
SQGVSSQSEETDINOLEKATQCTENPPEZGRPCRCVQDNPALYICROPHN
RFLICDCEBPFHGDVIGISBARKLLERBEDIYCNCTILYQVDETNGSATDOD
SGCRSVAGDGTCTSIPTIGTSESDQGIKRIEAPNSGKKKLTPOVPEAGAP
KCIQGVSSVAQPDPSVYCSNDCLLHAAATRRFLSGBRQKTRPKVTRKSLP
KCSYVQKISVHRLASREKRPVKKVLAASRSTGSKAASSTPSVSDHNYN
AVRKRPERKPTALSPTLISKCTPDKAFPGPSHRLGCLSRVRLVLAIVASS
SLPARKYQDASGPQVPLPSLMSLSGWLKSCVGLMLAIVSFRPM"

ORIGIN
Query Match 98.6%; Score 2828.2; DB 11; Length 4619;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2856; Conservative 0; Mismatches 8; Indels 3; Gaps 2;
1 TCCGTGTAGCTCTGGAATGCTGCGGATTCGCGCGCGGAGCTTTTAAAGG 60
68 TCCGTGTAGCTCTGGAATGCTGCGGATTCGCGCGCGGAGCTTTTAAAGG 127
61 CAGTCCCACTGTCCCTTGTGTGTTGAGAGCTCCGGAATCTTCAATTGATGACTG 120
128 CAGTCCCACTGTCCCTTGTGTGTTGAGAGCTCCGGAATCTTCAATTGATGACTG 187
121 ACTGCAACAGTGGGGTGAAGCTTGGCGGTGCTTGAACCTGAGCCCAAGTCTATATT 180
188 ACTGCAACAGTGGGGTGAAGCTTGGCGGTGCTTGAACCTGAGCCCAAGTCTATATT 247
181 TATGTAGATTTCCAGCCAAAGTTTCCAGTTTCAAGTTTGGACAGTATGAT 240
248 TATGTAGATTTCCAGCCAAAGTTTCCAGTTTCAAGTTTGGACAGTATGAT 307
241 AAGGGCACTGAGCAATGAGAGACCAACCAAGGCTATCAACCAACGTAAGAGTT 300
308 AAGGGCACTGAGCAATGAGAGACCAACCAAGGCTATCAACCAACGTAAGAGTT 367
301 AAGAAACCTGGGGTTTTCAGAAACCAAGATTCGCAACGTAAGAGTCCAGAGCAG 360
368 AAGAAACCTGGGGTTTTCAGAAACCAAGATTCGCAACGTAAGAGTCCAGAGCAG 427
361 GAGCGGACCCGAGTGAAGCAACCAAGCAATCTTCTGCGCGCGAGTGA 420
428 GAGTGAACCCGAGTGAAGCAACCAAGCAATCTTCTGCGCGCGAGTGA 487
421 GGGCAACCAACCTGAGAGAGGTTGAAGAGTTTCTTACACAGGTTGGCGCGAGG 480
488 GGGCAACCAACCTGAGAGAGGTTGAAGAGTTTCTTACACAGGTTGGCGCGAGG 547
481 AAAAAAATGTGCGCGGTCTCTGAGAGATTCAGTGAAGCCCAATCTTCAAGTCACT 540
548 AAAAAAATGTGCGCGGTCTCTGAGAGATTCAGTGAAGCCCAATCTTCAAGTCACT 607
541 GATGTGAGACAGCTTCCAGGGGAGCTTGAAGAGTTCAGATTCAGATTCAGATTC 600
608 GATGTGAGACAGCTTCCAGGGGAGCTTGAAGAGTTCAGATTCAGATTCAGATTC 667
601 GATGTGAGACAGCTTCCAGGGGAGCTTGAAGAGTTCAGATTCAGATTCAGATTC 660
668 GATGTGAGACAGCTTCCAGGGGAGCTTGAAGAGTTCAGATTCAGATTCAGATTC 727
661 GAGGAAGAAGACCTCTGACAGTGAACAGTGAAGCTTACGTTGAAGAACTTCAAGAC 720
728 GAGGAAGAAGACCTCTGACAGTGAACAGTGAAGCTTACGTTGAAGAACTTCAAGAC 787
721 CGCTTCGAGAAAGGAGAGCAAAACCTGTGAGAGGTCCTGAGAGGAGCTGAGAA 780
788 CGCTTCGAGAAAGGAGAGCAAAACCTGTGAGAGGTCCTGAGAGGAGCTGAGAA 847

QY 781 CGCTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 840
 DB 848 CGCTGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 907
 QY 841 AGTCCGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 900
 DB 908 AGTCCGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 967
 QY 901 TCCGAGTGAAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 960
 DB 968 TCCGAGTGAAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1027
 QY 961 ACAGAGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1020
 DB 1028 ACAGAGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1087
 QY 1021 AATGCCCTGATAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1080
 DB 1088 AATGCCCTGATAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1147
 QY 1081 CGGTGTGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1140
 DB 1148 CGGTGTGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1207
 QY 1141 CTGGAGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1200
 DB 1208 CTGGAGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1267
 QY 1201 GAGAGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1260
 DB 1268 GAGAGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1327
 QY 1261 GGCACAGCTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1320
 DB 1328 GGCACAGCTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1387
 QY 1321 AAGGTAAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1380
 DB 1388 AAGGTAAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1447
 QY 1381 CCTGTGTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1440
 DB 1448 CCTGTGTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1507
 QY 1441 CAGCTGAGCTGTGTGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1500
 DB 1508 CAGCTGAGCTGTGTGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1567
 QY 1501 AGATTTCTAAGTGAAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1560
 DB 1568 AGATTTCTAAGTGAAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1627
 QY 1561 CGAGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1620
 DB 1628 CGAGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1687
 QY 1621 CACAGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1680
 DB 1688 CACAGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1747
 QY 1681 AGGAGTGAAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1740
 DB 1748 AGGAGTGAAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1807
 QY 1741 GACCAAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1800
 DB 1808 GACCAAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1867
 QY 1801 ACCCAATGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1860
 DB 1868 ACCCAATGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1927
 QY 1861 CTGGGTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1920

DB 1928 CTGGGTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1987
 QY 1921 GCCAGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 1980
 DB 1988 GCCAGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2047
 QY 1981 CTGGGTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2040
 DB 2048 CTGGGTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2107
 QY 2041 GAGGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2100
 DB 2108 GAGGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2167
 QY 2101 TGGAGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2160
 DB 2168 TGGAGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2227
 QY 2161 CCAGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2220
 DB 2228 CCAGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2284
 QY 2221 CTGCTGATGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2280
 DB 2285 CTGCTGATGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2344
 QY 2281 CAATTTGATGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2340
 DB 2345 CAATTTGATGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2404
 QY 2341 ATCTGTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2400
 DB 2405 ATCTGTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2464
 QY 2401 TCATGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2460
 DB 2465 TCATGAGAGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2524
 QY 2461 TCAGTGTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2520
 DB 2525 TCAGTGTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2584
 QY 2521 CATTTGATGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2580
 DB 2585 CATTTGATGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2644
 QY 2581 GTGAATTTGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2640
 DB 2645 GTGAATTTGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2704
 QY 2641 CCAGATTTGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2700
 DB 2705 CCAGATTTGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2764
 QY 2701 TCTGTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2760
 DB 2765 TCTGTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2824
 QY 2761 GGGTGTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2820
 DB 2825 GGGTGTGAGAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2884
 QY 2821 GTTAAGTGAAGAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2867
 DB 2885 GTTAAGTGAAGAGAGAGAGAGATTTCTGCCGAACTGGGAGTGTCCAAATAGGC 2931

RESULT 2
 BC060249 4651 bp mRNA linear HTC 19-NOV-2003
 LOCUS BC060249 Mus musculus death inducer-oblierator 1, mRNA (cDNA clone
 DEFINITION IMAGE:6826043), containing frame-shift errors.

ACCESSION BC060249
 VERSION BC060249.1 GI:37805411
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 4651)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Scheefer, C.F., Bhat, N.K., Hopkins, R.P., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Staele, T.E., Soares, M.B., Bonaldi, M.P., Casavant, T.L., Schaefer, T.R., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carinici, P., Prange, C., Kahn, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huijck, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmitt, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalhus, D.B., Scherch, A., Schein, J.B., Jones, S.J. and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

TITLE 2 (bases 1 to 4651)
 Strausberg, R.
 Direct Submission
 Submitted (20-OCT-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
 NIH-MGC Project URL: http://mgc.ncl.nih.gov
 Contact: MGC help desk
 Email: cgapbs-rt@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library Preparation: M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 info@cgsc.bc.ca

COMMENT
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ram Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Matheson, Candice McNeely, Steven Ness, Pawan Pandon, Anna-Liisa Prabh, Parvaneh Saedi, Jacqueline Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Scott, Michael Thorne, Miranada Teal, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

FEATURES
 source
 1..4651
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:6826043"
 /tissue_type="Brain, mouse, 13.5, 14.5, 16.5, 17.5 dpc"
 /clone_id="NIH_BMAP_FY0"
 /lab_host="DH10B"

ORIGIN /note="Vector: pYX-Asc"

Query Match 96.8%; Score 2775.8; DB 11; Length 4651;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 2814; Conservative 0; Mismatches 7; Indels 4; Gaps 3;

44 GGAGCTGTTTAAAGGAGGAGTCCCACTGCTTGTGTGTTGAGAGTCCGGAATCTT 103
 90 GAGAGCTGTTTAAAGGAGGAGTCCCACTGCTTGTGTGTTGAGAGTCCGGAATCTT 149
 104 CTCATTGATGAACCTGACTGACCAACAGTGGGTGAGGCTTGTGCTTGTGACTGG 163
 150 CTCATTGATGAACCTGACTGACCAACAGTGGGTGAGGCTTGTGCTTGTGACTGG 209
 164 CCCAGGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 223
 210 CCCAGGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 269
 224 GAGCAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 283
 270 GAGCAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 329
 284 CCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 343
 330 CCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 389
 344 AGGTCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 403
 390 AGGTCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 449
 404 CCTGCGCGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 463
 450 CCTGCGCGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 509
 464 CGGTTGCG 523
 510 CGGTTGCG 569
 524 CATCTTCCACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 583
 570 CATCTTCCACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 629
 584 AGATCAGAAAGTGGCCCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 643
 630 AGATCAGAAAGTGGCCCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 689
 644 AGGCAAAAGAGGTGAAG 703
 690 AGGCAAAAGAGGTGAAG 749
 704 TGAAGGAACCTTCAAGACCCCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 763
 750 TGAAGGAACCTTCAAGACCCCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 809
 764 TGAAGGAACCTTCAAGACCCCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 823
 810 TGAAGGAACCTTCAAGACCCCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 869
 824 GAGAGTCCCAATATGAG 883
 870 GAGAGTCCCAATATGAG 929
 884 CTAGTCAAGGAGCCAGAGTCCAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 943
 930 CTAGTCAAGGAGCCAGAGTCCAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 989
 944 AGGCACTCAAGGAG 1003
 990 AGGCACTCAAGGAG 1049
 1004 GTGAGGTTTAAAGGAGGAGTCCCACTGCTTGTGTGTTGAGAGTCCGGAATCTT 1063

Db 1050 GTGAGTTTAAAGACCCCAATGCGCTGTATCTGATCTGCGCGCAGCTCAACAACAGGT 1109
Oy 1064 TTATGATCTGCTGATCGGTGTGAGAGAGTGTTCATAGTGACTGTGTGGTATTTCTG 1123
Db 1110 TTATGATCTGCTGATCGGTGTGAGAGAGTGTTCATAGTGACTGTGTGGTATTTCTG 1169
Oy 1124 AGGCGGAGAGGCGGCTCTGTGAAAGAAACGGGAAAGACTACATCTGCCAATTGGACCA 1183
Db 1170 AGGCGGAGAGGCGGCTCTGTGAAAGAAACGGGAAAGACTACATCTGCCAATTGGACCA 1229
Oy 1184 TTTTTCAGAGTSCAGAGATGAGCAAAACGTTAGCGCCCAATAGAGAGAGACTCTGGTGC 1243
Db 1230 TTTTTCAGAGTSCAGAGATGAGCAAAACGTTAGCGCCCAATAGAGAGAGACTCTGGTGC 1289
Oy 1244 GATCTGTGGGTGTGATGTGACAGACTGCAAGACATAGGAGACAGTATAGAGAGAGTCCG 1303
Db 1290 GATCTGTGGGTGTGATGTGACAGACTGCAAGACATAGGAGAGAGTATAGAGAGAGTCCG 1349
Oy 1304 GAGAAAGCCAGGAGCATTAAGGGGTAGATTGAGAAAGGACCAACCCAGCGCAAGAA 1363
Db 1350 GAGAAAGCCAGGAGCATTAAGGGGTAGATTGAGAAAGGACCAACCCAGCGCAAGAA 1409
Oy 1364 AACTCAAGATTTTTCAGGCTGTCTGTAGAGGCTCTGTGTGCTCTTAATGCAATGGCCCTG 1423
Db 1410 AACTCAAGATTTTTCAGGCTGTCTGTAGAGGCTCTGTGTGCTCTTAATGCAATGGCCCTG 1469
Oy 1424 GGTTGTTCAAGTGTAGACAGGCTGACTGTGTATTTGAGATTAATGACTGCAATCTCAAC 1483
Db 1470 GGTTGTTCAAGTGTAGACAGGCTGACTGTGTATTTGAGATTAATGACTGCAATCTCAAC 1529
Oy 1484 ACGGAGGAGCTACCAATGAGATTTCTAAGTTCAAGTTAAGAAAC-AAAAACAAACCCAG 1542
Db 1530 ACGGAGGAGCTACCAATGAGATTTCTAAGTTCAAGTTAAGAAACAAACAAACCCAG 1589
Oy 1543 GAAAGAGTCAAGAGAGAGAGGAAAGTTCAAGTCTTCCAAATGCAAGTTCAGGTGGG 1602
Db 1590 GAAAGAGTCAAGAGAGAGAGGAAAGTTCAAGTCTTCCAAATGCAAGTTCAGGTGGG 1649
Oy 1603 ATTAAATCTCTTCTGTGCAAGAGACTAGCGTCAAGAGAAAGGAAACCCAGTGAAG 1662
Db 1650 ATTAAATCTCTTCTGTGCAAGAGACTAGCGTCAAGAGAAAGGAAACCCAGTGAAG 1709
Oy 1663 AAGATGATCTGAGCTTCAGAGAGTGAATCTTGTGAAAGAGGACGCTGTGAGAGCAC 1722
Db 1710 AAGATGATCTGAGCTTCAGAGAGTGAATCTTGTGAAAGAGGACGCTGTGAGAGCAC 1769
Oy 1723 ACACCATCTGTGGGCAAGTGAACCAACTACATGCTGTGAAAGCCAGAGAGGCAAGAG 1782
Db 1770 ACACCATCTGTGGGCAAGTGAACCAACTACATGCTGTGAAAGCCAGAGAGGCAAGAG 1829
Oy 1783 CCACTGCACTCTGCGCCACACCTATTGATTAATGATAGTATCACCAAGGCTGGCTTC 1842
Db 1830 CCACTGCACTCTGCGCCACACCTATTGATTAATGATAGTATCACCAAGGCTGGCTTC 1889
Oy 1843 CCAAGGCCCCCTCCATCATCTGTGGTGTGCTGCGGCTGTCTAGGACCAAGATCTCGGT 1902
Db 1890 CCAAGGCCCCCTCCATCATCTGTGGTGTGCTGCGGCTGTCTAGGACCAAGATCTCGGT 1949
Oy 1903 GTTCTGTGTCTGATAGTGAACCAAGCTCACTGCAAGCCAGAGAGAGATACCAAGTGC 1962
Db 1950 GTTCTGTGTCTGATAGTGAACCAAGCTCACTGCAAGCCAGAGAGAGATACCAAGTGC 2009
Oy 1963 TCTGAGACCCAGGAGTCTGCTGTAGCCGTGTGAGAGCTCTCTGAGGAGTCTTAAGAGC 2022
Db 2010 TCTGAGACCCAGGAGTCTGCTGTAGCCGTGTGTAGAGCTCTCTGAGGAGTCTTAAGAGC 2069
Oy 2023 TGTGTAGGCTCTCATGTTGAGAGCAATTTCTTAATTCAGTTTATGAGCTGTGATGATTT 2082
Db 2070 TGTGTAGGCTCTCATGTTGAGAGCAATTTCTTAATTCAGTTTATGAGCTGTGATGATTT 2129
Oy 2083 GAACTGCTCTCAACAAGATGGAGCAACATCACTGCACTGTTCAGACTCAGTTCAAGCTTG 2142
Db 2130 GAACTGCTCTCAACAAGATGGAGCAACATCACTGCACTGTTCAGACTCAGTTCAAGCTTG 2189

Oy 2143 AGTTCTCAACAAGAGCCAGCAACAACCAATGAGATTTTCCCAAGTTACCTATTTTCT 2202
Db 2190 AGTTCTCAACAAGAGCCAGCAACAACCAAT- GAGTTTTCAGTTACCTATTTT--C 2246
Oy 2203 TCAACATGATGACCACTGCTCTCTGATAGAAAGATTTTCTTATGATAGGAGCAAT 2262
Db 2247 TTCAATGATGACCACTGCTCTCTCTATAGAAAGATTTTCTTATGATAGGAGCAAT 2306
Oy 2263 TTTGAAAGAGAGAGTGAACAATTTGTGTCTATTCTACATCTACACACTACATTAACG 2322
Db 2307 TTTGAAAGAGAGAGTGAACAATTTGTGTCTATTCTACATCTACACACTACATTAACG 2366
Oy 2323 GTTGAATTTGAAGATCATCTGCTCTCAAGATCCCATGATGATGATGATGATGATGAT 2382
Db 2367 GTTGAATTTGAAGATCATCTGCTCTCAAGATCCCATGATGATGATGATGATGATGAT 2426
Oy 2383 ACACCTACAGAGATTTCTCAACAAGAGATTTGAGGAGTGTAGCTTCCAGACTGCTGA 2442
Db 2427 ACACCTACAGAGATTTCTCAACAAGAGATTTGAGGAGTGTAGCTTCCAGACTGCTGA 2486
Oy 2443 GATGCCCCAGATTTCAAAATCAGTGTATTATGTTGATTAACCTTTGAAACCAAGATTA 2502
Db 2487 GATGCCCCAGATTTCAAAATCAGTGTATTATGTTGATTAACCTTTGAAACCAAGATTA 2546
Oy 2503 GATCACTGAGTTGAGGCAATTTCTGTTTCACTGCTGTACAAAGTACACCTTTTACTT 2562
Db 2547 GATCACTGAGTTGAGGCAATTTCTGTTTCACTGCTGTACAAAGTACACCTTTTACTT 2606
Oy 2553 TTTGACAGTGTGTTCAAGTGAATTTTCACTGCTGTACAAAGTACACCTTTTACTT 2622
Db 2607 TTTGACAGTGTGTTCAAGTGAATTTTCACTGCTGTACAAAGTACACCTTTTACTT 2666
Oy 2623 CTCTGTCCACCTCAGCCCAAGATTAAGGAGGCTCATGAGCTTAAGCTTAATTT 2682
Db 2667 CTCTGTCCACCTCAGCCCAAGATTAAGGAGGCTCATGAGCTTAAGCTTAATTT 2726
Oy 2683 CTGTAAGTTGGGGGTCTCTGCTGCCCCCACTGTAAGAGTCTGAACCTTAACAGGACA 2742
Db 2727 CTGTAAGTTGGGGGTCTCTGCTGCCCCCACTGTAAGAGTCTGAACCTTAACAGGACA 2786
Oy 2743 CACCTGAATGATCTGTGGGAGTGAACCTGAGAGTGTGAACAGAGTCAACAGCTGA 2802
Db 2787 CACCTGAATGATCTGTGGGAGTGAACCTGAGAGTGTGAACAGAGTCAACAGCTGA 2846
Oy 2803 GCTGCAATTTCAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2862
Db 2847 GCTGCAATTTCAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2906
Oy 2863 CTGGG 2867
Db 2907 CTGGG 2911

RESULT 3
AK014422 1614 bp mRNA linear HTC 20-SEP-2003
LOCUS AK014422
DEFINITION Mus musculus 18 days pregnant adult female placenta and extra embryonic tissue cDNA, RIKEN full-length enriched library, clone:3830408B01 product:death inducer-obliterator-1, full insert sequence.
ACCESSION AK014422
VERSION AK014422.1 GI:12852263
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253

PUBMED REFERENCE	10349636
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL MEDLINE PUBMED	Genome Res. 10 (10), 1617-1630 (2000)
REFERENCE	20499374
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Teshiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakauechi,S., Ikegami,T., Kasahiguchi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanabe,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
JOURNAL MEDLINE PUBMED	Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexillary sequencer
REFERENCE	Genome Res. 10 (11), 1757-1771 (2000)
AUTHORS	4 11076861
TITLE	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
JOURNAL REFERENCE	Functional annotation of a full-length mouse cDNA collection
AUTHORS	Nature 409, 685-690 (2001)
TITLE	5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
JOURNAL REFERENCE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
AUTHORS	Nature 420, 563-573 (2002)
TITLE	6 (bases 1 to 1614)
JOURNAL REFERENCE	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Horii,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kaubekawa,T., Kato,H., Kawai,Y., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyma,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sobabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toyota,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
JOURNAL REFERENCE	Direct Submission
AUTHORS	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsukuba-shi, Yokohama, Kanagawa 230-0045, Japan [E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel.81-45-503-9222, Fax:81-45-503-9216]
TITLE	COMMENT
JOURNAL REFERENCE	Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
AUTHORS	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science laboratory in RIKEN Division of Experimental Animal Research at Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGACATTCCTCAGTAAATTAAATTAATCCCCCCCCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 50.0. Second strand cDNA was prepared with the primer adpatent of sequence 15' GAGAGAGACATTCCTCAGTAAATTAAATTAATCCCCCCCCC 3'. cDNA was cleaved with BamHI and XhoI. Vector: a modified pluescript KS(+) after bulk excision from lambda PhiC I. Cloning sites, 5' end: SalI, 3' end: BamHI. Host: DH10B.
FEATURES	Location/Qualifiers
SOURCE	1..1614

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM,DB:3830408B01"
/db_xref="MG1:1910351"
/db_xref="taxon:10090"
/clone="3830408B01"
/sex="female"
/tissue_type="placenta and extra embryonic tissue"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev stage="18 days pregnant adult"
331..>1614
/note="death inducer-obiticator-1 (MGI:1344352,
GB|AJ238332, evidence: BLASTN, 99%, match=2599)
putative"

ORIGIN

Query Match      51.7%; Score 1482.4; DB 11; Length 1614;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1509; Conservative 0; Mismatches 6; Indels 3; Gaps 2
```

Query Match	Similarity	Score	1482.4	DB 11	Length	1614	
Best Local	Similarity	99.48	Pred	No. 0			
Matches	1509	Conservative	0	Mismatches	6	Indels	3
						Gaps	2
QY	1	TCCTGTAGACTCTGGAATGCTGCTGCGGATCCCGCGGCGCGGGAGCTTTGAAGAGG	60				
Db	100	TCCTGTAGACTCTGGAATGCTGCGGGATCCCGCGGCGCGGGAGCTTTGAAGAGG	155				
QY	61	CAGTCCCACTGTCCTCTTGTGTTTTCGAAGCTCCGGAAATCTTCTATTAATGAACGTG	120				
Db	160	CAGTCCCACTGTCCTCTTGTGTTTTCGAAGCTCCGGAAATCTTCTATTAATGAACGTG	215				
QY	121	ACTGCAACAGTGGGGTGAAGCTTGGGCGCTTGTGCTTGAACCTGGCCCAAGTCTTAAATTT	180				
Db	220	ACTGCAACAGTGGGGTGAAGCTTGGGCGCTTGTGCTTGAACCTGGCCCAAGTCTTAAATTT	275				
QY	181	TATGTAGATTTCAAGCCAAAGGTTTCAAGCTTTCAAGTTTGGACAGGTATGATGAT	240				
Db	280	TATGTAGATTTCAAGCCAAAGGTTTCAAGCTTTCAAGTTTGGACAGGTATGATGAT	339				
QY	241	AAAGGGCACTTGACGAATGAGAACCAACCAAGGCTATCAAAACCAAGTAAGAGTTTC	300				
Db	340	AAAGGGCACTTGACGAATGAGAACCAACCAAGGCTATCAAAACCAAGTAAGAGTTTC	399				
QY	301	AGGAATACTGGGGGTTTTCGAAGAACCAAGATTGCCAAACGTGAGGGTGCAGAGACAG	360				
Db	400	AGGAATACTGGGGGTTTTCGAAGAACCAAGATTGCCAAACGTGAGGGTGCAGAGACAG	459				
QY	361	GAGCGGACCCCGATGAGCAGCAACCAAGCAGCATTAACCTTCTCCCTGCGCGCAGTGA	420				
Db	460	GAGCGGACCCCGATGAGCAGCAACCAAGCAGCATTAACCTTCTCCCTGCGCGCAGTGA	519				
QY	421	CGGCAACCAAAACGTACGTAGAGGGGTGAAGAGTTTCTTACACAGGTTGGCGGCCGAGGG	480				
Db	520	CGGCAACCAAAACGTACGTAGAGGGGTGAAGAGTTTCTTACACAGGTTGGCGGCCGAGGG	579				
QY	481	AAAAAGATGTGCGCGGTGCTCCCTGAGAGATTCCAGTGAAGCCACATCTTTCACAGTCACT	540				
Db	580	AAAAAGATGTGCGCGGTGCTCCCTGAGAGATTCCAGTGAAGCCACATCTTTCACAGTCACT	639				
QY	541	GATGTGAGACACAGCTTCCGAGGGGAGCTTTGAAGACAGTTTTCAGAGTTCAGAAATGCGCCT	600				
Db	640	GATGTGAGACACAGCTTCCGAGGGGAGCTTTGAAGACAGTTTTCAGAGTTCAGAAATGCGCCT	699				
QY	601	GTAATGTACTCCTTGAAGGAAAGAACATCTGCTCTTTCGAAGAAAGCAAAAGAGGTGA	660				
Db	699	GTAATGTACTCCTTGAAGGAAAGAACATCTGCTCTTTCGAAGAAAGCAAAAGAGGTGA	756				
QY	661	GAGGAAGAGACACCTGTGACAGTGAACGTGATGCGCTTACGTTGAAGAACTTCAAGAC	720				
Db	757	GAGGAAGAGACACCTGTGACAGTGAACGTGATGCGCTTACGTTGAAGAACTTCAAGAC	816				
QY	721	CGCCTTCGAGAAACCGAGAGCAACACTGTGAGAGGCTCCTTGAGAGGCACTCGAAT	780				
Db	817	CGCCTTCGAGAAACCGAGAGCAACACTGTGAGAGGCTCCTTGAGAGGCACTCGAAT	876				

```

QY 781 CGCCCTGAGAGAGAGAGAGAGAGAGATTCCTGCCGAACTGGAGTGTCCAAATAGGC 840
DB 877 CGCCTGAGAGAGAGAGAGAGAGAGATTCCTGCCGAACTGGAGTGTCCAAATAGGC 936
QY 841 AGTCCGAGAGAGAGAGAGAGAGAGATTCCTGTAAAGAGAGAGAGAGAGAGAGAGAG 900
DB 937 AGTCCGAGAGAGAGAGAGAGAGAGATTCCTGTAAAGAGAGAGAGAGAGAGAGAGAG 996
QY 901 TCCGAGTCAAGAGAGAGAGAGAGAGAGATTCAGTTGAGAGAGAGAGAGAGAGAGAG 960
DB 997 TCCGAGTCAAGAGAGAGAGAGAGAGAGATTCAGTTGAGAGAGAGAGAGAGAGAGAG 1056
QY 961 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 1057 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1116
QY 1021 AATGCCCCCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 1080
DB 1117 AATGCCCCCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 1176
QY 1081 CGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
DB 1177 CGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1236
QY 1141 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1237 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1296
QY 1201 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1297 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1356
QY 1261 GGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 1357 GGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1416
QY 1321 AAGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1417 AAGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476
QY 1381 CCTGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 1477 CCTGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1536
QY 1441 CAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
DB 1537 CAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1596
QY 1501 AGATTTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1518
DB 1597 AGATTTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1614

```

```

RESULT 4
LOCUS BC029110
DEFINITION Mus musculus death inducer-oblierator 1, mRNA (cDNA clone
IMAGE:3980050), containing frame-shift errors.
ACCESSION BC029110
VERSION BC029110.1 GI:22137689
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1640)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, K.H., Shennan, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Djakshenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stepleton, M., Soares, M.B., Bonald, M.F., Casavant, T.L.,
Schneitz, T.B., Brownstein, M.J., Udell, T.B., Toshimaki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Miliush, S.J., Bosak, S.A., McSwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hilyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Pahay, J., Helton, E., Kerteman, M., Madan, A., Young, A.C., Shevchenko, Y.,
Sanchez, A., Whiting, M., Madan, A., Touchman, J.W., Green, E.D.,
Boutard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Gilwood, J., Schmitz, J., Myers, R.M.,
Butcherfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.B.,
Scherer, A., Schein, J.B., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 1640)
Strausberg, R.
Direct Submission
Submitted (01-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LUNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hilyk, S.W., Louised, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LUNL at: http://image.llnl.gov
Series: IRAX Plate: 23 Row: K Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein
This clone has the following problem: frame shifted.
Location/Qualifiers
1..1640
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/c1one="IMAGE:3980050"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha
model. 10 month old virgin mouse. Taken by biopsy."
/c1one_lib="NCI_CGAP_Mam1"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"

```

ORIGIN

```

Query Match 48.8% Score 1398.8; DB 11; Length 1640;
Best Local Similarity 93.0%; Pred. No. 0;
Matches 1522; Conservative 0; Mismatches 2; Indels 112; Gaps 1;

```

```

Db 121 GGGGTGAGGCTTGGCCGCTGCTGTCAGCTCGGACCCCAAGGCTATAATTTATGATGATTT 180
Qy 193 CCAGCCAAAGGTTTCCAGCTTTGAGTTTGGAC----- 227
Db 181 CCAGCCAAAGGTTTCCAGCTTTGAGTTTGGACAGGGTTTGTGTTTATGCGAGTCT 240
Qy 228 ----- 227
Db 241 CCCCCTTAAGTGAACATGACAGTAACTTTCATTACTAGCATTTAATGCAAGTTAAAG 300
Qy 228 -----AGTATGATGATTAAGGCGACTGAGCAATGA 260
Db 301 CCTGAGACGACGAGAAAAGTATCTCAGGATATGATATGAAGGCGACTGAGCAATGA 360
Qy 261 GGAAGCACTCCAGGCTATCAAAACCCACGATTAAGAGTTTCAGGAAACCTGGGGTTTTCG 320
Db 361 GGAAGCACTCCAGGCTATCAAAACCCACGATTAAGAGTTTCAGGAAACCTGGGGTTTTCG 420
Qy 321 AAGAACAAGATTTGCGCAAACTGAGAGGTGACAGAGACACGAGGCGGACCCCAAGTGA 380
Db 421 AAGAACAAGATTTGCGCAAACTGAGAGGTGACAGAGACACGAGGCGGACCCCAAGTGA 480
Qy 381 GCAACCAACGACGAGATTAACCTCTCCCTGCGCGCAGTGAACGCGCAACCAAAAGTACTGA 440
Db 481 GCAACCAACGACGAGATTAACCTCTCCCTGCGCGCAGTGAACGCGCAACCAAAAGTACTGA 540
Qy 441 GAGGCTGAGAAAGTTTCTTACACGCTTCGCGCGCGGAGGAGAAAGAAATGTCGGCTTC 500
Db 541 GAGGCTGAGAAAGTTTCTTACACGCTTCGCGCGCGGAGGAGAAAGAAATGTCGGCTTC 600
Qy 501 CCTGAGAGATTCAGTGAAGCCCACTCTTCCACAGTCACTGATGAGAGACAGCTTCCGA 560
Db 601 CCTGAGAGATTCAGTGAAGCCCACTCTTCCACAGTCACTGATGAGAGACAGCTTCCGA 660
Qy 561 GGGGAGCTTGAAGCACTTCTGAGATCAAGATGAGGCGCTGATCTGACTTTCAGGAA 620
Db 661 GGGGAGCTTGAAGCACTTCTGAGATCAAGATGAGGCGCTGATCTGACTTTCAGGAA 720
Qy 621 AGAATCTCTGCTCTTCTTCTGAAAAGGCAAAAGAGGTGAAGGAGAAAGACCTCTGA 680
Db 721 AGAATCTCTGCTCTTCTTCTGAAAAGGCAAAAGAGGTGAAGGAGAAAGACCTCTGA 780
Qy 681 CAGTGAACGATGAGGCTTCTTCTGAAAAGGCACTTCTGAAAAGGCACTTCTGAAAAGGCA 740
Db 781 CAGTGAACGATGAGGCTTCTTCTGAAAAGGCACTTCTGAAAAGGCACTTCTGAAAAGGCA 840
Qy 741 GCAAGAACCTGTGAGAGGCTCTGAGAGGCACTGAGAAATGCTGAGAGAGAGCGCAG 800
Db 841 GCAAGAACCTGTGAGAGGCTCTGAGAGGCACTGAGAAATGCTGAGAGAGAGAGCGCAG 900
Qy 801 AGAGGAATTTCTGCGCAAACTGAGAGGTGCCAAATAGGAGTGCAGAGCAAGAGC 860
Db 901 AGAGGAATTTCTGCGCAAACTGAGAGGTGCCAAATAGGAGTGCAGAGCAAGAGC 960
Qy 861 TCTCTGTAAGCAGAGGCTGAGAGGCTAGTCAAGGACCAAGTGTCCAGTCAAGAGCAGATGA 920
Db 961 TCTCTGTAAGCAGAGGCTGAGAGGCTAGTCAAGGACCAAGTGTCCAGTCAAGAGCAGATGA 1020
Qy 921 CATAGAAAATCAGTTGAAAGGAGAGGCACTCAGGAGAAATACAGAGAGAAACCCCAAGGA 980
Db 1021 CATAGAAAATCAGTTGAAAGGAGAGGCACTCAGGAGAAATACAGAGAGAAACCCCAAGGA 1080
Qy 981 AGCGGGCAAAACCAAGGCTGAGAGGTGAGAGTTTAAAGCCCAATGCTGATCTGATCTG 1040
Db 1081 AGCGGGCAAAACCAAGGCTGAGAGGTGAGAGTTTAAAGCCCAATGCTGATCTGATCTG 1140
Qy 1041 CCGCAGGCTCTCAACAACAGGTTTATGATCTGCTGATGCTGATGAGAGTGTGTTCA 1100
Db 1141 CCGCAGGCTCTCAACAACAGGTTTATGATCTGCTGATGCTGATGAGAGTGTGTTCA 1200
Qy 1101 TGGTGAATGTTGTTGTTTCTTGAAGCCCGAGGCGGCTCTCTGAGAAAGAAAGGAGGA 1160
Db 1201 TGGTGAATGTTGTTGTTTCTTGAAGCCCGAGGCGGCTCTCTGAGAAAGAAAGGAGGA 1260

```

```

Qy 1161 CTACATCTGCCCAATATTCACCAATTTTTCAGAGTGCAGAGTGAAGACAAACGTTAGCCAC 1220
Db 1261 CTACATCTGCCCAATATTCACCAATTTTTCAGAGTGCAGAGTGAAGACAAACGTTAGCCAC 1320
Qy 1221 CAATGACAGAGACTCTGGGTGAGATCTGTGTGGTCTGATATGCGACAGACTGCAAGACT 1280
Db 1321 CAATGACAGAGACTCTGGGTGAGATCTGTGTGGTCTGATATGCGACAGACTGCAAGACT 1380
Qy 1281 AGGACAGTGAAGAGAGAGTCCGAGAGAGACAGGCGCATTAAGGTTAGATTAAGAGGC 1340
Db 1381 AGGACAGTGAAGAGAGAGTCCGAGAGAGACAGGCGCATTAAGGTTAGATTAAGAGGC 1440
Qy 1341 AGCAAAACCCAGCGGCAAGAAAACCTCAGATATTTTCAGGCTGTCGATGAGGCTCTTG 1400
Db 1441 AGCAAAACCCAGCGGCAAGAAAACCTCAGATATTTTCAGGCTGTCGATGAGGCTCTTG 1500
Qy 1401 TGCTCTTAATGATGATGCTGCTGGGTGTTTCCAGTATGACAGAGCTGATCTGTGTATG 1460
Db 1501 TGCTCTTAATGATGATGCTGCTGGGTGTTTCCAGTATGACAGAGCTGATCTGTGTATG 1560
Qy 1461 CAGTATGATGATGATGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1520
Db 1561 CAGTATGATGATGATGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
Qy 1521 AGAACAACCAAAACCAAAA 1536
Db 1621 AGAACAACCAAAACCAAAA 1636

RESULT 5
BC014615
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCES
AUTHORS
TITLE
JOURNAL
REMARK

```

BC014615 2752 bp mRNA linear HTC 17-DEC-2003

Homo sapiens cDNA clone IMAGE:3532298, containing frame-shift errors.

BC014615

BC014615.2 GI:33869935

HTC.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 2752)

Strausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G., Klausner, R. D., Collins, F. S., Wagner, L., Shenmen, C. M., Schuler, G. D., Altschul, S. F., Zeeberg, B., Buetow, K. H., Scheffer, C. F., Bhat, N. K., Hopkins, R. F., Jordan, H., Moore, T., Max, S. I., Wang, J., Hsieh, L., Diatchenko, L., Marusina, K., Farmer, A. A., Rubin, G. M., Hong, L., Stapleton, M., Soares, M. B., Bonaldo, M. F., Casavant, T. L., Scheetz, T. E., Brownstein, M. J., Umedin, T. B., Toshimiyuki, S., Carninci, P., Prange, C., Raha, S. S., Loquellano, N. A., Peters, G. J., Abramson, R. D., Mullahy, S. J., Bosak, S. A., McEwan, P. J., McKernan, R. J., Malek, J. A., Gunaratne, P. H., Richards, S., Worley, K. C., Hale, S., Garcia, A. M., Gay, L. J., Bulik, S. W., Villalón, D. K., Muzny, D. M., Sodergren, E. J., Lu, X., Gibbs, R. A., Fahey, J., Heltan, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A. C., Shevchenko, Y., Bouffard, G. G., Blakeley, R. W., Touchman, J. W., Green, E. D., Dickinson, M. C., Rodriguez, A. C., Grimwood, J., Schmutz, J., Myers, R. M., Butlerfield, Y. S., Krzywicki, M. I., Skalska, U., Smalhe, D. B., Scherf, A., Schein, J. E., Jones, S. J., and Marra, M. A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 2752)

Strausberg, R.

Direct Submission

Submitted (24-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT

On Aug 19, 2003 this sequence version replaced gi:15779097.
Contact: MGC help desk
Email: cgaab8-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systembiology.org
contact: amadan@systembiology.org
Anup Madan, Jessica Pahey, Brin Helton, Mark Ketteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAL Plate: 11 Row: 3 Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 18375616
This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
1..2752
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGB:353228"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_id="NIH-MGC-17"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

ORIGIN

Query Match 35.3%; Score 1010.8; DB 11; Length 2752;
Best Local Similarity 71.8%; Pred. No. 3.2e-236;
Matches 1481; Conservative 0; Mismatches 532; Indels 51; Gaps 10;
16 GAAATGGCTGCGGAGTCCCGGCGCGGAGAGCTTTTAAGAGGAGTCCCACTGTC 75
1 GAAATGGCTGCGGAGTCCCGGCGCGGAGAGCTTTTAAGAGGAGTCCCACTGAGT 60
76 CTTTGGTGTTCGAAAGCTCCGGAATCTTCATTGATGAACTGTGACTGACCAAGTGG 135
61 AATCGAGTGTGAAAAATACAGAGCTTTACTCGTGAACAGTGTGACTGACCAAGTGT 120
136 GT-GAGGCTTGCGGCTGCTTG-----ACCTGCGCCAGGCTTAATAATTT 181
121 GTTGAAGCGCTGCTGTCTGTGACCCAGAGTTCGTCTGCGAGGCTTTTGTGTGT 180
182 ATGTAGATTTCCAGCCAAAGTTTCCAGCTTTCAGTTTGGGAAGAGTAATGATGATA 241
181 ATTATGATTT-CAGGAAAAAGTCCAAAGCTTTCAGTGTGAGCAAGTATGACGACA 239
242 AAGGCACTGAGCAATGAGAAAGCAAGGCTTCAAAACCCACAGTAAGAGTTCA 301
240 AAGGAGCCCGAGCAATGAGAGGCACTTAAGGCCATCAAAACCAAGCAAGAGTTCA 299
302 GAAAAAAGCTGGGCTTTCGAAAGAACAGATTCGCAAGTGAAGGTGACAGACAG 361
300 GAAAAAAGTGGGGTTTTCGAAAGAACCACTATGCGCAAGGAGGCGGAGGAGCGCG 359
382 AAGCCGAGCCCGAGTGAAGCAAC-----ACAGAGCATTAACCTCTCCCTGCGCGCA 415
360 AAGCTGACCCCACTGAGAGCGGCAACCCCAAGAGAGAGTGGGCTGTCTCCGCGCGCA 419
416 GTTGAAGGCAACCAAAAGTACTGAGAGGAGTGAAGAGTTTCTTCAACAGTTTCGCGCC 475
420 GTTGAAGGCAAGCCCAAGGCACTGAGCGGCTGAGAGAGTTCTGACATTTGGCGCGCGC 479
476 GAGGAAAAAGAAATGTGCGGCTGTCTTGAAGAGATTCAGATGAGCCCAATCTTCAAG 535
480 GGGGAGGAGGAGAGCTGTCTCTCTGAGAGATTTCTGTGAGCCCAAGTCTCTGCGCGC 539
536 TCACGATGTGAGACAGTTCGAGAGGAGGAGTGAAGAGTTCTGAGATTCAGAGAGT 595
540 CCACAGAGCCGAGACAGCTCCGAGAGGAGCGGTGAAGAGCGCTTTCGAGACCAAGAGCG 599

QY 596 GCCCTGATCTGACTCTTA---GGAAAAACATCTGCTCTTCTGAAAAAGCAAAAG 652
DB 600 GCCCCAGTCTGCTTCACAGCTGTGAAAGAAACGACAGCTCTCTGAAAAAGTGAAG 659
QY 653 GAGGTGAAGAAAGAAAGACCTTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 712
DB 660 GAGGAGATTAACAGATGACACTTCGATGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAG 719
QY 713 TTCAAGACCGCTTCGAGAAAG 772
DB 720 TTCAAGATGCTTCGAG 779
QY 773 GTCAAGATGCTTCGAGAAAG 832
DB 780 TCAGAGTGGCTTCGAGAAAG 839
QY 833 AATATGAGAGTCCGAG 892
DB 840 AAGCCAGTGAACATGTGAG 899
QY 893 GACAGATGCTCCAGTGAAG 952
DB 900 GGGTGTGCTCCAGAGCTGGAGAAAGATGACAGAGAGAGTAAAGTTGAGAGAGAGAGAGAG 959
QY 953 AGGAAATTAAGAGAGAAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1012
DB 960 AGGACATCAAGATGAG 1019
QY 1013 AGAGCCCAAG 1072
DB 1020 AGAGCCCAAG 1079
QY 1073 GCTGTATGCTGTGAG 1132
DB 1080 GCTGTATGCTGTGAG 1139
QY 1133 GCGGCTCTTGAAG 1192
DB 1140 GAGAGCTTTGAGAAAG 1199
QY 1193 TCGAGATGAGCAAAAG 1252
DB 1200 TCGAGATGAGCACTTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1259
QY 1253 GTGCTATGAGCAAG 1312
DB 1260 ATGCTATGAGCAAG 1319
QY 1313 AGGCAATTAAG 1372
DB 1320 AAGGATTAAG 1379
QY 1373 TATTCAGAGCTGTGAG 1432
DB 1380 TCTTCAGAGCTGTGAG 1439
QY 1433 GTTGAAGCAAG 1492
DB 1440 AGTGGCGAG 1499
QY 1493 GTACATGAGATTTCTAAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1552
DB 1500 CGACATGAGATTTCTAAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1559
QY 1553 AG 1612
DB 1560 AGATGAG 1619
QY 1613 CTTTGTGAG 1668
DB 1620 CTTTGTGAG 1679

QY 1669 ATGCTGCTTCAGAGTGAAGTCTTCTGGAGAGGAGGAGCCTGTGAGAGCAGACACCA 1728
 DB 1680 GTGTGCTCCGCGGAGTGAAGCACTCGGAGAGGAGGAGGCTTTGAGAGGAGCAGCGCG 1739
 QY 1729 TCTGGGAGAGTGAAGCACTTATGAGTCTGTGAAGCCAGAGAGCCAGAGAGCCCACT 1788
 DB 1740 TGTGGGAGAGTGAAGCACTTATGAGTCTGTGAAGCCAGAGAGGAGAGGAGCTG-----CT 1790
 QY 1789 GCACTCTGCCCCAGCCTTATGAGTCTGTGAAGTCTGTGAAGCCAGAGAGGCTTCCAGGC 1848
 DB 1791 GCTCCCTGCGCGTCACTGTGTGAAGTCTGTGAAGTCTGTGAAGGCTTCCCTGTGAGC 1850
 QY 1849 CCTCCCTGAGTCACT-----GGGTGCTGCTGGGCTGTCTGAGCAG-----AGTCT 1898
 DB 1851 CCTCCCTGCTTCTTCTGTGAGTCACTCCCTGAGCCTGTCTGAGAGCTGGAGTTCAGCT 1910
 QY 1899 GAGTGTCTGTGCTGATAGTGAAGCAGAGC-----TCACTGAGCCAGAGAGCAGATACC 1954
 DB 1911 TTGTGTGAAGTGAATCAAGACAGACCGGCTGACATGAGCGGAGAGCAGAGCCATGTCC 1970
 QY 1955 AAGATGCTCTGAGACCCAGGCTGTCTGCTGAGTCTGTGAGCCTCTCTGTGGGTGCTCC 2014
 DB 1971 AGGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2030
 QY 2015 TAAAGAGCTGTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2038
 DB 2031 TCCTAAGCTGTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2054

RESULT 6
 AK044919 3385 bp mRNA linear HTC 20-SEP-2003
 DEFINITION Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
 induced library, clone:B13001M08 product:death
 inducer-oblierator-1, full insert sequence.
 AK044919
 VERSION AK044919.1 GI:26336914
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS Carrinci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

REFERENCE
 AUTHORS Carrinci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 JOURNAL prepare full-length cDNA libraries for rapid discovery of new genes
 MEDLINE 20499374
 PUBMED 11042159

TITLE
 JOURNAL Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carrinci, P.,
 Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system-384-format
 JOURNAL sequencing pipeline with 384 multiplexed sequencer
 MEDLINE Genome Res. 10 (11), 1757-1771 (2000)
 PUBMED 20530913
 REFERENCE 11076861
 AUTHORS 4
 TITLE The RIKEN Genome Exploration Research Group Phase II Team and the
 PANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5
 AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 JOURNAL of 60,770 full-length cDNAs
 MEDLINE Nature 420, 563-573 (2002)
 PUBMED 6 (bases 1 to 3385)
 REFERENCE
 AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carrinci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hirakawa, T., Hirozane, T.,
 Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kondo, H., Kouda, M.,
 Koya, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohse, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takaishi, F., Takaku-Akanita, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-JUL-2001) Yoshinori Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 COMMENT
 CDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Dr. Tomohiro Kono (Department of Animal
 Science, Tokyo University of Agriculture, 1377 Hnako Atsugi City,
 Kanagawa Prefecture, Japan) whose assistance we gratefully
 acknowledge.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/.
 FEATURES
 source location/Qualifiers
 1..3385
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM DB:B13001M08"
 /db_xref="MGI:2410273"
 /db_xref="taxon:10090"
 /clone="B13001M08"
 /tissue_type="parthenogenote"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="9.5 days embryo"
 2443..3276
 /note="unlabeled protein product; death
 inducer-oblierator-1 (MGI:1344352, GB|A238332,
 evidence: BLASTN, 99%, match=2599)
 putative"
 /codon_start=1
 /protein_id="BAC32141.1"
 /db_xref="GI:26336915"
 /translation="MDKGLSNBAPNAIKPTSKFRFTWGFRTTAKRBAQDTE
 VPSKQPOOHLSLRSGRQPKTERVEFLTYARRCKKVPVSLBDSSEPTSTV
 TDVETASBSVSSEIRSGPDSIGKHPASBRKAGSEEDTSDSDSLTLKE
 LONRLRRKRGQPVPRSLRGSONRLPKRRERDSAPTSVOIGSAQDRPLCKQBPBA
 SGPVQSSTDIENQLBKAQTQATERNPRAGKPRPCRYVDNALYICICRPHNN
 R"

ORIGIN
 Query Match 29.1%; Score 835.6; DB 11; Length 3385;
 Best Local Similarity 98.4%; Pred. No. 2.6e-193;
 Matches 844; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
 QY 211 GCTTTCAGTTTGGAGCAGATGATGAATGAAGGACCTGAGCAATAGAGACACCC 2710


```

Db      2422 GATTGCTTTTGTCTAGTATGATATGATTAAGGACCTTACGACATATGAGAGACCTC 2481
Qy      271  AAGGCTATCAAAACCCACAGTAAGAGTTTCAAGAAAACCTGGGGTTTTCAGAAACACAG 330
Db      2482 AAGGCTATCAAAACCCACAGTAAGAGTTTCAAGAAAACCTGGGGTTTTCAGAAACACAG 2541
Qy      331  ATTGCAAAACGTGAGGGTTCAGAGAGACAGAGGCGGACCCAGTGAAGACAAACACAG 390
Db      2542 ATTGCAAAACGTGAGGGTTCAGAGAGACAGAGGCGGACCCAGTGAAGACAAACACAG 2601
Qy      391  CAGCAATACCTTCCCTGCGCGCGAGTGAACCGCAACCAAAACCTACTGAGAGGGTGA 450
Db      2602 CAGCAATACCTTCCCTGCGCGCGAGTGAACCGCAACCAAAACCTACTGAGAGGGTGA 2661
Qy      451  GAGTTTCTTACACAGGTTGCGGCGCGAGGAGAAAAGATGTCGCGGTTCCTGAGAGAT 510
Db      2662 GAGTTTCTTACACAGGTTGCGGCGCGAGGAGAAAAGATGTCGCGGTTCCTGAGAGAT 2721
Qy      511  TCCAGTGAAGCCCAATCTTCCACAGTCACTGATGTGAGAGACAGCTTCCGAGGGAGCGTT 570
Db      2722 TCCAGTGAAGCCCAATCTTCCACAGTCACTGATGTGAGAGACAGCTTCCGAGGGAGCGTT 2781
Qy      571  GAAACAGATTCTGAGATCAGAAATGCGCCCTGTATCTGACTCTCTTAGGAGAAACATCTCT 630
Db      2782 GAAACAGATTCTGAGATCAGAAATGCGCCCTGTATCTGACTCTCTTAGGAGAAACATCTCT 2841
Qy      631  GCCTCTTCTGAAAAGGCAAAAGAGGTGAAGAGAAAGACACCTCTGACAGTGAAGT 690
Db      2842 GCCTCTTCTGAAAAGGCAAAAGAGGTGAAGAGAAAGACACCTCTGACAGTGAAGT 2901
Qy      691  GATGACCTTACCTGTGAAGAGACTTCAGAACCGCTTCGAGAAAGCGAGAGCAAGACCT 750
Db      2902 GATGACCTTACCTGTGAAGAGACTTCAGAACCGCTTCGAGAAAGCGAGAGCAAGACCT 2961
Qy      751  GTGAGAGAGTCCCTGAGAGGCTGACAGATTCGCTCTGAGAGAAAGCGCAGAGAGAGAT 810
Db      2962 GTGAGAGAGTCCCTGAGAGGCTGACAGATTCGCTCTGAGAGAAAGCGCAGAGAGAGAT 3021
Qy      811  TCTGCGAAACCTGAGAGTCCCAATATGAGCACTGCGAGACAGACAGACCTCTCTTAAG 870
Db      3022 TCTGCGAAACCTGAGAGTCCCAATATGAGCACTGCGAGACAGACAGACCTCTCTTAAG 3081
Qy      871  CAGAGACCTGAGGCTAGTCAAGGAGCCAGTGTCCCACTGAGAGACAGATGATGAAAT 930
Db      3082 CAGAGACCTGAGGCTAGTCAAGGAGCCAGTGTCCCACTGAGAGACAGATGATGAAAT 3141
Qy      931  CAGTTGAAGGAGAGCGGCACTCAAGGAAATACAGAGAAAACCCAGGAGAGCGGCAAA 990
Db      3142 CAGTTGAAGGAGAGCGGCACTCAAGGAAATACAGAGAAAACCCAGGAGAGCGGCAAA 3201
Qy      991  CCAAGGCTGAGTGAAGGTTTACAGACCCCAATGCGCTGTACTGAGTCTGCGGACAGCT 1050
Db      3202 CCAAGGCTGAGTGAAGGTTTACAGACCCCAATGCGCTGTACTGAGTCTGCGGACAGCT 3261
Qy      1051 CACAACAACAGGTTTATG 1068
Db      3262 CACAACAACAGGTTATG 3279

```

```

RESULT 7
LOCUS   BUS22670      874 bp      mRNA      linear      EST 13-SEP-2002
DEFINITION  AGRCOURT 10158094 NCI_CGAP Co24 Mus musculus cDNA clone
IMAGE:552926 5', mRNA sequence.
ACCESSION  BUS22670
VERSION    BUS22670.1 GI:22830196
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus; Chordata; Vertebrata; Euteleostomi;
            Eumetazoa; Metazoa; Chordata; Rodentia; Sclerozoa; Muriidae; Mus.
REFERENCE  1 (bases 1 to 874)

```

```

AUTHORS  NIH-MGC http://mgc.nci.nih.gov/.
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished (1999)
COMMENT  Contact: Robert Strausberg, Ph.D.
          Email: cgaps-remail.nih.gov
          Tissue Procurement: The Cepko Laboratory
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNL at:
          http://image.lnl.gov
          Plate: LLM14129 row: a column: 02
          High quality sequence stop: 720.
          Location/Qualifiers
            1..874
              /organism="Mus musculus"
              /mol_type="mRNA"
              /strain="FVB/N"
              /db_xref="taxon:10090"
              /clone="IMAG:552926"
              /lab_host="DH10B (T1 phage-resistant)"
              /clone_1ib="NCI CGAP Co24"
              /note="Organ: colon; Vector: pCMV-Sport6; Site 1: NotI;
              Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
              Average insert size 1.6 kb. Constructed by Life
              Technologies. Note: this is a NCI CGAP library."

```

ORIGIN

```

Query Match      27.2%  Score 780.8; DB 13; Length 874;
Best Local Similarity 99.5%; Pred. No. 66-180; 2; Indels 2; Gaps 2;
Matches 804; Conservative 0; Mismatches 2;

Qy      797  GCGAGAGAGAGATTTCTGCCGAACCTGGAGTGTCCAAATAGGACAGTCCGAGCAGACA 856
Db      1  GCGAGAGAGAGATTTCTGCCGAACCTGGAGTGTCCAAATAGGACAGTCCGAGCAGACA 60
Qy      857  GACCTCTCTGTAAGCAGAGGCTGAGGCTAGTCAAGGACCAAGTGTCCAGTCAAGACAG 916
Db      61  GACCTCTCTGTAAGCAGAGGCTGAGGCTAGTCAAGGACCAAGTGTCCAGTCAAGACAG 120
Qy      917  ATGACATAGAAATCACTGTTGAAGGAGGACGACTCAGGAAATACAGAGGAAACCCCA 976
Db      121  ATGACATAGAAATCACTGTTGAAGGAGGACGACTCAGGAAATACAGAGGAAACCCCA 180
Qy      977  GCGAGCGGCGCAACCAAGGCTGAGTGAAGTTTACAGACCCCAATGCGCTGTACTGCA 1036
Db      181  GCGAGCGGCGCAACCAAGGCTGAGTGAAGTTTACAGACCCCAATGCGCTGTACTGCA 240
Qy      1037  TCTGCCGACAGCTCAACAACAAGTTTATGATCTGTGTATGATGATGATGATGATGAT 1096
Db      241  TCTGCCGACAGCTCAACAACAAGTTTATGATCTGTGTATGATGATGATGATGATGAT 300
Qy      1097  TCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1156
Db      301  TCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Qy      1157  AAGACTACATGTCGCCCAATTTGCAACATTTTCAAGTGAAGTGAAGTGAAGTGAAGT 1216
Db      361  AAGACTACATGTCGCCCAATTTGCAACATTTTCAAGTGAAGTGAAGTGAAGTGAAGT 420
Qy      1217  CCAACAATGAGAGAGACTGTGGGTGAGATCTGTGGTCTGTATGAGCACTGCAACA 1276
Db      421  CCAACAATGAGAGAGACTGTGGGTGAGATCTGTGGTCTGTATGAGCACTGCAACA 480
Qy      1277  GCATAGGAGCAGTGAAGCAGAGTCCGAGGAGACCAAGGCGCTAAAGGATGAGATTGAG 1336
Db      481  GCATAGGAGCAGTGAAGCAGAGTCCGAGGAGACCAAGGCGCTAAAGGATGAGATTGAG 540
Qy      1337  AGGAGCAAAACCCAGCGGCAAGAAAACCTCAAGATATTTCAAGCTGTGTGAGAGCTC 1396
Db      541  AGGAGCAAAACCCAGCGGCAAGAAAACCTCAAGATATTTCAAGCTGTGTGAGAGCTC 600

```

QY 1397 CTGTCCTCTTAATGACATGCGCTGGGTGTTCCAGTGTACAGACGCTGCTGTGT 1456
 DB 601 CTGGTCTCTTAATGACATGCGCTGGGTGTTCCAGTGTACAGACGCTGCTGTGT 660
 QY 1457 ATTGAGTAATGACATGCTATTCATCAACAGCAGACGCTACATGAGATTTCTAAGTTGAG 1516
 DB 661 ATTGAGTAATGACATGCTATTCATCAACAGCAGACGCTACATGAGATTTCTAAGTTGAG 720
 QY 1517 GTAAAGAAC-AAAAAACAACCCCAAGAAAGTTCAGAGACGACGAAAGTTGAGT 1575
 DB 721 GTAAAGAACAAAAAACAACCCCAAGAAAGTTCAGAGACGACGAAAGTTGAGT 780
 QY 1576 CTTCCAAATGAC-AGTGTTCAGTGGGG 1602
 DB 781 CTTCCAAATGACAGTGTTCAGTGGGG 808
 RESULT 8
 LOCUS BI078113 796 bp mRNA linear EST 20-JUN-2001
 DEFINITION 6028724201 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5004317 5',
 mRNA sequence.
 ACCESSION BI078113
 VERSION BI078113.1 GI:14496443
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 796)
 NIH-MGC http://mgs.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rtmail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLN at:
 http://image.llnl.gov
 Plate: LLM11042 row: 9 column: 06
 High quality sequence stop: 742.
 Location/Qualifiers
 1. 796
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N-3"
 /db_xref="taxon:10090"
 /clone="IMAGE:5004317"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="5 months"
 /lab_host="DH10B"
 /clone_id="NCI_CGAP_Mam2"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"
 ORIGIN
 Query Match 24.5%; Score 702.2; DB 12; Length 796;
 Best Local Similarity 98.1%; Pred. No. 1e-160; Indels 7; Gaps 7;
 Matches 784; Conservative 0; Mismatches 8; Indels 7; Gaps 7;
 QY 1439 CACAGCCTGACTCTGTGATTCAGATGATGATGCTGCTCAACACGAGAGCTTACCA 1498
 DB 1 CACAGCCTGACTCTGTGATTCAGATGATGATGCTGCTCAACACGAGAGCTTACCA 60
 QY 1499 TCAGATTTCTAAGTTCAAGTAAAGAACAAAAAACAACCCCAAGAAAGTTCAGAGCA 1558
 DB 61 TCAGATTTCTAAGTTCAAGTAAAGAACAAAAAACAACCCCAAGAAAGTTCAGAGCA 120

QY 1559 AGCCAGAA-AAAGTCAGTCTTCCAAAATGACAGTGTTCAGTGGAGTTAAATCTCTCT 1617
 DB 121 AGCCAGAAAGTTCAGTCTTCCAAAATGACAGTGTTCAGTGGAGTTAAATCTCTCT 180
 QY 1618 GTGCACAAAGAGCTAGCGCTCAGAGAAAGGAAAAACCAAGTGAAGAAAGTGTGCTGCT 1677
 DB 181 GTGCACAAAGAGCTAGCGCTCAGAGAAAGGAAAAACCAAGTGAAGAAAGTGTGCTGCT 240
 QY 1678 TCCAGAGATGAGACTTCTGAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1737
 DB 241 TCCAGAGATGAGACTTCTGAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
 QY 1738 AGTGACCAACTACTCAATGCTGTGAAGCAGAGAAAGCCAGAGAAAGCCACTGCTCTG 1797
 DB 301 AGTGACCAACTACTCAATGCTGTGAAGCAGAGAAAGCCAGAGAAAGCCACTGCTCTG 360
 QY 1798 CCCACCTTATGAGTAAATGTAAGTATCAACCAAGAGCTGCTTCCAGGCCCCCTCCAT 1857
 DB 361 CCCACCTTATGAGTAAATGTAAGTATCAACCAAGAGCTGCTTCCAGGCCCCCTCCAT 420
 QY 1858 CATCTGGGTGGCTGCTGGGCTGTCTAGAGACAGAGTCTGGGTGTTCTGGTGGATTA 1917
 DB 421 CATCTGGGTGGCTGCTGGGCTGTCTAGAGACAGAGTCTGGGTGTTCTGGTGGATTA 480
 QY 1918 GTAGCAGAGAGCTCACTGCGACAGAGACAGATACCAAGATGCTCTGGA-CCCCAGGT 1976
 DB 481 GTAGCAGAGAGCTCACTGCGACAGAGACAGATACCAAGATGCTCTGGA-CCCCAGGT 540
 QY 1977 GTTCTGCTTACGCTGTGAGAGCTCTCTGGGTGTTCTTAAAGCTGTGAGGCTCAT 2036
 DB 541 GTTCTGCTTACGCTGTGAGAGCTCTCTGGGTGTTCTTAAAGCTGTGAGGCTCAT 600
 QY 2037 GTTGAAGGCAATTTCTTATTTAGATTTAGAGCTTGTGATTTGAAC-TGCTCTC-A 2094
 DB 601 GTTGAAGGCAATTTCTTATTTAGATTTAGAGCTTGTGATTTGAAC-TGCTCTC-A 660
 QY 2095 ACAAGATGGAACAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2154
 DB 661 ACAAGATGGAACAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
 QY 2155 GACAGCCAGACAAACCAATGAGATTTTCCAGTTACTTATTTTCTCAACATGCTCA 2214
 DB 721 GACAG-CCAGACAAACCAT-GAGTTTCCCAAGTTACTTATTTCTTC-ACATGCTCA 777
 QY 2215 CCACTGCTTCTTATTAAGAA 2233
 DB 778 CCACTGCTTCTTATTAAGAA 796
 RESULT 9
 LOCUS BY713603 992 bp mRNA linear EST 17-DEC-2002
 DEFINITION BY713603 RIKEN full-length enriched, 18 days pregnant, placenta and
 extra embryonic tissue Mus musculus cDNA clone 3830408B01 5', mRNA
 sequence.
 ACCESSION BY713603
 VERSION BY713603.1 GI:27125612
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 992)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaide, I., Otao, N., Saito, R., Suzuki, H., Yamada, I.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
 Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapath, A., Matsuda, H.,
 Batalov, S., Beisel, K.W., Blake, J.A., Brdtk, D., Brusic, V.,
 Chochia, C., Corbani, J.E., Cousins, S., Dalla, B., Dragani, T.A.,
 Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, B.D., Kanai, A.,

ACCESSION CF169555 musculus cDNA clone NIA:B0815C09 IMAGE:30468896 5', mRNA sequence.
 VERSION CF169555.1 GI:33279104
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 667)
 Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
 Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method
 Genome Res. 11 (9), 1553-1558 (2001)
 JOURNAL 21429098
 MEDLINE 11544199
 PUBMED
 COMMENT Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@igsun.grc.nia.nih.gov
 Plate: B0815 row: C column: 09
 Seq primer: M13 Reverse
 High quality sequence stop: 667
 POLYA=No.

FEATURES
 source
 1..667
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /db_xref="taxon:10090"
 /clone="NIA:B0815C09 IMAGE:30468896"
 /dev_stage="Newborn Kidney"
 /lab_host="DH10B"
 /clone_1="NIA Mouse Newborn Kidney cDNA Library (long 1)"
 /note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). In brief, double-stranded cDNAs were synthesized with an oligo(dT) primer (Invitrogen): 5'-pAGCTAGTCTAGATCGCAGCGCGCCCTTTTCTTTT-3' from 26 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L-SalI, purified by phenol/chloroform, and separated from free linkers by Centricion 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and centrifuged 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.0 Kb. The library was constructed by Yulan Piao."

ORIGIN
 Query Match 23 2%; Score 665.4; DB 14; Length 667;
 Best Local Similarity 99.9%; Pred. No. 1e-151; Indels 0; Gaps 0;
 Matches 666; Conservative 0; Mismatches 1;
 628 CCTGCTCTTTGAAAGCAAAAGAGGTGAAGAGAAAGACACTTCTGACAGTGAAC 687
 1 CCTGCTCTTTGAAAGCAAAAGAGGTGAAGAGAAAGACACTTCTGACAGTGAAC 60
 688 AGTGAATGCTTGAAGTGAAGCACTTCAAGACCGCTTGGAGAAAGGAGAGCAAGAA 747
 61 AGTGAATGCTTGAAGTGAAGCACTTCAAGACCGCTTGGAGAAAGGAGAGCAAGAA 120
 748 CCTGGAAGAGGTCTCTGAGAGGCACTCAAGATCGCTTGAAGAAAGGAGAGAGAA 807

Db 121 CCTGTGAGAGAGTCCCTGAGAGGACAGTCAAGATCCGCTTGAGAGAAAGGAGAGAGAA 180
 Oy 808 GATTCTGCCGAAACTGGAGAGTCCAAATAGGACAGTCCGACAGAGACAGCTCTCT 867
 Db 181 GATTCTGCCGAAACTGGAGAGTCCAAATAGGACAGTCCGACAGAGACAGCTCTCT 240
 Oy 868 AAGCAGAGAGCTGAGGCTAGTCAAGGACAGAGTCCCAATCAGAGACAGATGACATGAA 927
 Db 241 AAGCAGAGAGCTGAGGCTAGTCAAGGACAGAGTCCCAATCAGAGACAGATGACATGAA 300
 Oy 928 AATCAGTTGAAAGGAGAGGAGCTCAAGGAAATATACAGAGAAACCCCAAGAGAGGAG 987
 Db 301 AATCAGTTGAAAGGAGAGGAGCTCAAGGAAATATACAGAGAAACCCCAAGAGAGGAG 360
 Oy 988 AAACCAAGCTGAGTGTGAGGTTTACGACCCCAATCCCTGTATCGCATGCGCCGAC 1047
 Db 361 AAACCAAGCTGAGTGTGAGGTTTACGACCCCAATCCCTGTATCGCATGCGCCGAC 420
 Oy 1048 CCTCAACAACAGAGTTTATGATCTGCTGATCGGTGAGAGAGTGTCCATGCTGAC 1107
 Db 421 CCTCAACAACAGAGTTTATGATCTGCTGATCGGTGAGAGAGTGTCCATGCTGAC 480
 Oy 1108 TGTGTGGATATTTCTGAGGCGGAGGCGGCTCTCTGAAAGGAGAGGAGACTATC 1167
 Db 481 TGTGTGGATATTTCTGAGGCGGAGGCGGCTCTCTGAAAGGAGAGGAGACTATC 540
 Oy 1168 TGCCCAATTTGACCACTTTTGAAGTGCAGATGAGACAAAGGTAGCGGCACATGAG 1227
 Db 541 TGCCCAATTTGACCACTTTTGAAGTGCAGATGAGACAAAGGTAGCGGCACATGAG 600
 Oy 1228 CAGAGCTCTGGGTCAGATCTGTGGGTGCTGATGAGCAGACACTGACAAAGCATAGGACA 1287
 Db 601 CAGAGCTCTGGGTCAGATCTGTGGGTGCTGATGAGCAGACACTGACAAAGCATAGGACA 660
 Oy 1288 GTAGAGC 1294
 Db 661 GTAGAGC 667

RESULT 11 691 bp mRNA linear EST 26-OCT-2001
 BB637982 RIKEN full-length enriched, 3 days neonate thymus Mus
 LOCUS BB637982.1 GI:16473713
 DEFINITION musculus cDNA clone A630095A07 5', mRNA sequence.
 ACCESSION BB637982
 VERSION BB637982.1
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 691)
 Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Komano, H., Konda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
 Unpublished (2001)
 TITLE JOURNAL
 COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenihiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-resgsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komano, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to

size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCAGCAGC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

ORIGIN

Query Match 21.5%; Score 617.4; DB 13; Length 632;
 Best Local Similarity 99.4%; Pred. No. 5.9e-140;
 Matches 629; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY	13	CTGGAATAGCTGCGGATCCCGGCGCGGAGCTTTTAAAGAGCAGTCCCACTG	72
DB	1	CTGGAATAGCTGCGGATCCCGGCGCGGAGCTTTTAAAGAGCAGTCCCACTG	60
QY	73	TTCCCTTTGGTGTTCGAAGCTCCGGAATCTTCTCATTAAGAACTGACAGCAAGT	132
DB	61	TTCCCTTTGGTGTTCGAAGCTCCGGAATCTTCTCATTAAGAACTGACAGCAAGT	120
QY	133	GGGGTGAAGCTTGGCCGCTCTGCTGACCTGGCCCAAGCTTAATTTATGTAGGATT	192
DB	121	GGGGTGAAGCTTGGCCGCTCTGCTGACCTGGCCCAAGCTTAATTTATGTAGGATT	180
QY	193	CCAGCCAAAGTTTTCAGAGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCACTG	252
DB	181	CCAGCCAAAGTTTTCAGAGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCACTG	240
QY	253	AGCAATGAGAAAGCCCAAGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCACTG	312
DB	241	AGCAATGAGAAAGCCCAAGCTTTCAGTTTGGAGACAGTATGATGATTAAGGCACTG	300
QY	313	GGTTTTCAGAAACACAGATTCAGAAAGTTCAGAGGTCAGAGACAGGAGCGGACCC	372
DB	301	GGTTTTCAGAAACACAGATTCAGAAAGTTCAGAGGTCAGAGACAGGAGCGGACCC	360
QY	373	AGTGAAGACCAACAGAGCATTAACCTCTCCCTGCGCGGAGTGAAGCGCAACCAAA	432
DB	361	AGTGAAGACCAACAGAGCATTAACCTCTCCCTGCGCGGAGTGAAGCGCAACCAAA	420
QY	433	CGTACTGAGAGGTTCAGAAAGTTCATTAACAGGTTTCGGCCCGGAGGAAAGATG	492
DB	421	CGTACTGAGAGGTTCAGAAAGTTCATTAACAGGTTTCGGCCCGGAGGAAAGATG	480
QY	493	CCGGTCTCCCTGAGGATTCAGATGAGCCCATCTTCACAGTCACTGATGAGACA	552
DB	481	CCGGTCTCCCTGAGGATTCAGATGAGCCCATCTTCACAGTCACTGATGAGACA	540
QY	553	GCTTCCAGAGGAGGAGTTGAAAGAGTTCAGATGAGAGTGGCCCTGATCTGACTCC	612
DB	541	GCTTCCAGAGGAGGAGTTGAAAGAGTTCAGATGAGAGTGGCCCTGATCTGACTCC	600
QY	613	TTAGGAAAGAAACATCTCTCTCTCTCTGAAAG	645
DB	601	TTAGGAAAGAAACATCTCTCTCTCTCTGAAAG	632

RESULT 13
 CFS35718 623 bp mRNA linear EST 12-SEP-2003
 LOCUS UI-M-GH0-cg2-h-20-0-UI.r1 NIH BMAP GH0 Mus musculus cDNA clone
 DEFINITION IMAGE:30534619 5', mRNA sequence.
 ACCESSION CFS35718
 VERSION CFS35718.1 GI:34587686
 KEYWORDS EST.
 ORGANISM Mus musculus (house mouse)
 SOURCE Mus musculus
 REFERENCE 1 (bases 1 to 623)

AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: <http://genome.uiowa.edu/distribution/mouse1.html>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

COMMENT

seq primer: pX-5.

FEATURES

Location/Qualifiers
 1..623
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:30534619"
 /tissue_type="whole brain"
 /dev_stage="1, 5, and 15 days newborn"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH BMAP GH0"
 /note="Organ: Brain; Vector: pX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCAGCAGT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

ORIGIN

Query Match 21.2%; Score 607.4; DB 14; Length 623;
 Best Local Similarity 99.4%; Pred. No. 1.7e-137;
 Matches 619; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY	302	GGAAACCTGGGGTTTTCAGAAACCAAGATTCAGAAAGTTCAGAGAGACACGG	361
DB	1	GGAAACCTGGGGTTTTCAGAAACCAAGATTCAGAAAGTTCAGAGAGACACGG	60
QY	362	AGGCGAGCCCGAGT-GAGGAGAACCAAGAGCATTAACCTCTCCCTGCGCGGAGTGA	420
DB	61	AGGCGAGCCCGAGT-GAGGAGAACCAAGAGCATTAACCTCTCCCTGCGCGGAGTGA	120
QY	421	CGGCAACCAAAAGTTCAGAGGTCAGAAAGTTCATTAACAGGTTTCGGCCCGGAGGG	480
DB	121	CGGCAACCAAAAGTTCAGAGGTCAGAAAGTTCATTAACAGGTTTCGGCCCGGAGGG	180
QY	481	AAAAAGATGTCGCGGTCTCCCTGAGAGATTCAGTGAAGCCCATCTTCCACAGTACT	540
DB	181	AAAAAGATGTCGCGGTCTCCCTGAGAGATTCAGTGAAGCCCATCTTCCACAGTACT	240
QY	541	GATGAGAGAGCTTCGAGAGGAGCGTTGAAAGAGTTCAGATGAGAGTGGCCCT	600
DB	241	GATGAGAGAGCTTCGAGAGGAGCGTTGAAAGAGTTCAGATGAGAGTGGCCCT	300
QY	601	GATCTGACTCTTTCAGGAAAGAAACATCTCTCTCTTGAAGAGCAAGAGGTGA	660
DB	301	GATCTGACTCTTTCAGGAAAGAAACATCTCTCTCTTGAAGAGCAAGAGGTGA	360
QY	661	GAGGAAGAAACACTTCAGACAGTGAAGTGGCTTAACTTGAAGAACTTCAGAAC	720
DB	361	GAGGAAGAAACACTTCAGACAGTGAAGTGGCTTAACTTGAAGAACTTCAGAAC	420

OY		721	CGCCTTGGGAGAAAGCCAGAGACAGAAGAACCTGTGGAGAGGTGCCCTGAGAGGCAGTCAGAT	780
Dd		421	CGCCTTGAGAAAAGCCAGAGACAGAAGAACCTGTGGAGAGGTGCCCTGAGAGGCAGTCAGAT	480
OY		781	CGCCTGAGGAAGAAAGCCAGAGAGGAAGAAATTCTGCCGAACAATCGGAGAGTGTCCAATAAGGC	840
Dd		481	CGCCTGAGGAAGAAAGCCAGAGAGGAAGAAATTCTGCCGAACAATCGGAGAGTGTCCAATAAGGC	540
OY		841	AGTCCCGAGCAGACAGACACTCTCTGTAAAGCAGAGAGCCTGAGAGCTAATGTCAGAGCACTAGT	900
Dd		541	AGTCCCGAGCAGCAGACACTCTCTGTAAAGCAGAGAGCCTGAGAGCTAATGTCAGAGCACTAGT	600
OY		901	TCCCAGTCAGAGCAGATGATCAT	923
Dd		601	TCNCACTGAGAGCAGATGATCAT	623
<hr/>				
RESULT 14				
BB636983				
LOCUS BB636983				
DEFINITION musculus cDNA clone A530047115 5', mRNA sequence.				
ACCESSION BB636983				
VERSION BB636983.1 GI:16472769				
KEYWORDS EST.				
SOURCE Mus musculus (house mouse)				
ORGANISM Mus musculus				
REFERENCE Mue Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS Arakawa,T., Carinci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Komno,H., Koude,M., Koye,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okasaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Saeki,D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.				
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)				
TITLE Unpublished (2001)				
JOURNAL Contact: Yoshitake Hayashizaki				
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel.: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-ree@sec.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Carinci,P., Shibata,Y., Hayasu,N., Sugahara,Y., Shibata,K., Itoh,M., Komno,H., Okasaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) vegli,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Isikawa,T., Ozawa,K., Tanaka,T., Maenuma,S., Kawai,J., Okasaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multichannel sequencer. Genome Res. 10 (11), 1757-1771 (2000) Komno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carinci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Kondo,S., Shingawa,A., Saito,T., Kiyosawa,H., Yamanake,I., Akawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y. Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001) Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.				

	FEATURES	e mouse tissues.
	Location/Qualifiers	
	source	
	1..645	
	/organism="Mus musculus"	
	/mol_type="mRNA"	
	/db_xref="taxon:10090"	
	/clone="A530047115"	
	/sex="male"	
	/tissue_type="aorta and vein"	
	/dev_stage="adult"	
	/lab_host="DH10B"	
	/clone_1lb="RIKEN full-length enriched, adult male aorta and vein"	
	/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'] GAGAGAGAAGATCCACAGCTCTTTTCTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAATTTCGAGTTATTAATAATATCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLX I."	
ORIGIN		
Query Match	20.7%; Score 594.6;	DB 10; Length 645;
Best Local Similarity	99.2%; Pred. No. 2.3e-134;	
Matches 608; Conservative	0; Mismatches 4; Indels 1; Gaps 1;	
Gy	1 TCGGTGTAGCTCTGGAAATGGCTGCGGAATCCGCGGCGCGGAGACTGTATTAAGG	60
Dd	34 TCCGTGTAGCTCTGGAAATGGCTGCGGAATCCGCGGCGCGGAGACTGTATTAAGG	93
Gy	61 CAGTCCCACATGTCCTTTGGTTGTGTGAAGCTCCGATCTTCTCATATGAATCTGTG	120
Dd	94 CAGTCCCACATGTCCTTTGGTTGTGTGAAGCTCCGATCTTCTCATATGAATCTGTG	153
Gy	121 ACTGACAACAGTGGGGTGAGGGCTTGCGCCTTCTGCTTGAACCTGGCCCAAGCTCTAATTT	180
Dd	154 ACTGACAACAGTGGGGTGAGGGCTTGCGCCTTCTGCTTGAACCTGGCCCAAGCTCTAATTT	213
Gy	181 TATGTAGATTTCCAGGCCAAGGTTTCCAAGCTTTCAGTTTGTGGACAGATATGATGAT	240
Dd	214 TATGTAGATTTCCAGGCCAAGGTTTCCAAGCTTTCAGTTTGTGGACAGATATGATGAT	273
Gy	241 AAAAGGCACCCTGAGCATATGAGGAAGCACCCAGGCTATCAAACCCACAGTAAGGATTC	300
Dd	274 AAAAGGCACCCTGAGCATATGAGGAAGCACCCAGGCTATCAAACCCACAGTAAGGATTC	333
Gy	301 AGAAAACTTGGGTTTTTCGAAGAACCAAGATTCCCAAACGTGAAGGTGACAGAGACAG	360
Dd	334 AGAAAACTTGGGTTTTTCGAAGAACCAAGATTCCCAAACGTGAAGGTGACAGAGACAG	393
Gy	361 GAGGCGGACCCCATGAGACACCAACAGACATTAACCTCTCCCTGCGCCGACAGTGA	420
Dd	394 GAGGTGACCCCATGAGACACCAACAGACATTAACCTCTCCCTGCGCCGACAGTGA	453
Gy	421 CGGCAACAAAAGGTATGAGAGGGTGAAGAGTTTCTTATCAAGGTTCCGCGCCGAGGG	480
Dd	454 CGGCAACAAAAGGTATGAGAGGGTGAAGAGTTTCTTATCAAGGTTCCGCGCCGAGGG	513
Gy	481 AAAAAAGATGTGCGGTGTCCCTGAGAGATTCAGTAGGCCACATCTTCCACAGTCACT	540
Dd	514 AAAAAAGATGTGCGGTGTCCCTGAGAGATTCAGTAGGCCACAT--TTTCAAGTCACT	572
Gy	541 GATGTGAGACAGCTTCCAGGGGAGCGCTGAAAAGCAGTTCTAGATTCAGAAAGTGCCCT	600

Db 317 AGGAAACCTGGGTTTCGAAAGACAGATGCCAAACGTAGAGGTCCAGAGACAG 376
QY 361 GAGGCGAGCCCACTGAGCAGCAACCAAGCAATTAACCTCTCCCTGCGCCGAGTGA 420
Db 377 GAGGTGACCCCACTGAGCAGCAACCAAGCAATTAACCTCTCCCTGCGCCGAGTGA 436
QY 421 CGGCAACCAAAACCTGAGAGAGGTAGAGAGTTTCTTACCAAGGTTCCGCGCCGAGG 480
Db 437 CGGCAACCAAAACCTGAGAGAGGTAGAGAGTTTCTTACCAAGGTTCCGCGCCGAGG 496
QY 481 AAAAGATGTCGCGGTGTCCTGAGAGATTCAGTGAAGCCCAATCTTCCAGTCACT 540
Db 497 AAAAGATGTCGCGGTGTCCTGAGAGATTCAGTGAAGCCCAATCTTCCAGTCACT 556
QY 541 GATGTGAGACGCTTCGAGGGAGCGTTGAAAGCAAGTTCTGAGATCAGAAAGTGCCT 600
Db 557 GATGTGAGACGCTTCGAGGGAGCG-TGAAAGCAAGTTCTGAAATCAAAAGGCGCCT 615
QY 601 GTATCTGACTCTTATGGAAGAAACATCTGCTCTCTTGAAGCAAGCAAGGAGGTGA 660
Db 616 GT-TCTTACTCTTATGGAAGAA-AAAATCTGCTCTTGAAGCAAGGAGGAGGAAA 673
QY 661 GAGGAAGAGACACTCTGACAGTGAAGATGAGCTTACG 702
Db 674 GAANNAAAACCTTTNAGNGAGAGAGGCTTACCTTAAG 715

Search completed: April 23, 2004, 18:03:08
Job time : 7086.26 secs

Matches 169; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

QY 662 GAAAAGTGAAGAGGAGATGACCACTTCCGATGATGACAGGATGCTG 741
DB 886 GACGGGGAGACGGGGAGACGAGAGACGGGGAGAGACGAGAGACGGGGAG 827
QY 742 ACCTTGAAGAGCTTCAAGATCGCTTCCGAGAGAGCGGAGACAGAGCCCACTGAGAG 801
DB 826 GACGGGGAGAGAGACGAGAGACGGGGAGAGACGAGAGAGAGAGAGAGAGAG 767
QY 802 CCCCTGAAGAGATCAGAGTCCCTGCGAGAGAAAGCCCGGAGAGAGGCTCCGCGAG 861
DB 766 GACGAGAGACGGGGAGAGACGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 707
QY 862 ACTGTGGGCTCCGAGGCGCAGTGAACACTGTGAGAGGCGCTTCCGATGAGAGAGCC 921
DB 706 GACGGGGAG 647
QY 922 GAGAACGATCAGGGGGGTTGTCTCCAGAGCTGGGAGAAAGATGACAGAGAGTAAAGTTGAG 981
DB 646 GACGAGAGACGGGGAGAGACGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 587
QY 982 GAAAGGCGGCTCAGAGACATCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1041
DB 586 GCGGAGAGACGGGGAG 527
QY 1042 GATGTGAGGCTTACGA 1058
DB 526 GACGGGAGAGAGAGCA 510

RESULT 7
US-08-781-891-208/c
; Sequence 208, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-Ken
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSER: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; SOFTWARE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 208:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16442 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-08-781-891-208

Query Match 1.6%; Score 42.6; DB 3; Length 16442;
Best Local Similarity 48.5%; Pred. No. 0.24;
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 774 GAAAGGGAACAGAGAGCCCACTGAGAGAGCCCTGAAAGGATCCAGAGTCCGCTGCGAA 833
DB 16430 GAG 16371
QY 834 GAAAGGCGGAGAGAGAGGCTCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 893
DB 16370 GAG 16311
QY 894 GAGGCTCTGCTCCCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 953
DB 16310 GAG 16251
QY 954 GAAAGATGACAGAGAGATGAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1013
DB 16250 GAG 16191
QY 1014 G 1014
DB 16190 G 16190

RESULT 8
US-09-618-166-208/c
; Sequence 208, Application US/09618166
; Patent No. 6583112
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-Ken
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Seed Intellectual Property Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; SOFTWARE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/618,166
; FILING DATE: 17-Jul-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: McMaister, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 240052.419c1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 208:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16442 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 208:
US-09-618-166-208

Query Match 1.6%; Score 42.6; DB 4; Length 16442;

Best Local Similarity 48.5%; Pred. No. 0.24;
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 774 GAACCGGAAACGAGAGCCCACTGAGAGCCCTTGAAGGATCCGAGATCCGCTGGCGAA 833
Db 16430 GAGCAG 16371
QY 834 GAAGCGCCGGAG 893
Db 16370 GCAAG 16311
QY 894 GAGCGCTCTCCCACTGAG 953
Db 16310 GAG 16251
QY 954 GAAGATGACAG 1013
Db 16250 GAG 16191
QY 1014 G 1014
Db 16190 G 16190

RESULT 9
US-08-867-030B-5

Sequence 5, Application us/0867030B
Patent No. 5948900
GENERAL INFORMATION:
APPLICANT: Yoher et al.
TITLE OF INVENTION: Streptococcus pneumoniae
TITLE OF INVENTION: Capsular Polysaccharide Genes and Flanking Regions
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Benjamin Aaron Adler, Ph.D., J.D.
STREET: 8011 Candle Lane
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch floppy
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word for Macintosh
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,030B
FILING DATE: June 2, 1997
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: 08/243,546
FILING DATE: May 16, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Benjamin Aaron Adler, Ph.D., J.D.
REGISTRATION NUMBER: 35,423
REFERENCE/DOCKET NUMBER: D5923
TELECOMMUNICATION INFORMATION:
TELEPHONE: (713) 777-2321
TELEFAX: (713) 777-6908
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4951 bp
TYPE: nucleic acid
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: other nucleic acid
HYPOTHEICAL: no
ANTY-SENSE: no
FRAGMENT TYPE:
ORIGINAL SOURCE:
IMMEDIATE SOURCE:
POSITION IN GENOME:

US-08-867-030B-5

Query Match 1.6%; Score 41.8; DB 2; Length 4951;
Best Local Similarity 49.8%; Pred. No. 0.19;
Matches 106; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
QY 2237 ACAACAGCCGTGTATACAGCTTCAGTTTCCAGATTGTCACGCGCAGACATGAAAGCA 2296
Db 1238 AGMAAACTCTATACACCGATTATTTTGGCAGAGAAATAGGGAAATATTTTATATGA 1297
QY 2297 TCACTCTTTTATGTTTGGGAATCTTTCAGATGATGTTGATCATCTGATTTTCAAGT 2356
Db 1238 TACATTATTTTAACTGTTTGGATTTTTCAGATCATGATTTTCACTTTATGTT 1357
QY 2357 GTACATTTATTTTACTGCGCAGATATGGGATTTTTCATATGCCATTGAC 2416
Db 1358 GTTTTTCCTTATCTTATCTGTTGGCGGTTTATATTTTCAATGCTGATGATATA 1417
QY 2417 AGCTTACACCCACCATGACATTCGACTT 2449
Db 1418 GTCTTACAGTTGTAGTATGATGAGAGATT 1450

RESULT 10

PCT-US95-06119-5
Sequence 5, Application PC/TUS9506119
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE CAPSULAR
TITLE OF INVENTION: POLYSACCHARIDE GENES AND FLANKING REGIONS
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
SOFTWARE: Patent Release #1.0, Version
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06119
FILING DATE: CONCURRENTLY HERewith
CLASSIFICATION:
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: US 08/243,546
FILING DATE: 16-MAY-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: AMCT018P--
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4951 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: other nucleic acid
PCT-US95-06119-5

2237 ACAACAGCGCTGTAATACGCTTCAGTTTTCACATTTGTCACAGCCAGACATGAAAGA 2296
1238 AGAAAACTATATACACCGGATTTATTTGGACAGAAATAGGGAAATATTTTATATGA 1297
2297 TCACCTCTTTTATGTTGGGAATCTTGCAGTATGTTGACATGATTTTCAAGT 2356
1298 TACATTTATTTTAAATGTTGGATTTTTCAGATCATGATTTTCAATTCCTTATGTT 1357
2357 GTACATTTATTTTACGCGGAGATAGGGATTTTTCATTCATGTCGATTCAC 2416
1358 GTTTTTCCTTATTTATTCATTCGTTGGCGGCTTATATTTTCATGCTGATATATA 1417
2417 AGCTACACACCCACATACATTCGACTT 2449
1418 GTCTACAGTTGATGTAAGTATGAGAGATT 1450

RESULT 11

US-09-710-794-3
Sequence 3, Application US/09710794
Patent No. 6573069
GENERAL INFORMATION:
APPLICANT: Holloway, James L.
APPLICANT: Gao, Zeren
APPLICANT: Whitmore, Theodore E.
TITLE OF INVENTION: NOVEL CRIB PROTEIN ZMS1
FILE REFERENCE: 99-76
CURRENT APPLICATION NUMBER: US/09/710,794
CURRENT FILING DATE: 2000-11-09
PRIORITY FILING DATE: 1999-11-10
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 1068
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Degenerate polynucleotide sequence for human zms1
NAME/KEY: m1sc_feature
LOCATION: (1)...(1068)
OTHER INFORMATION: n = A,T,C or G
US-09-710-794-3

Query Match 1.6%; Score 41.4; DB 4; Length 1068;
Best Local Similarity 22.4%; Pred. No. 0.086;
Matches 102; Conservative 79; Mismatches 275; Indels 0; Gaps 0;

513 CCGCGGCGAGAGACATGCTGTCTCCCTGAGAGATTCGGTGAAGCCACGCTCGGCC 572
102 YMGCAVATCNATGCAVGTNGMNGCNGANGVGNITTYGAGATACWMSNTTYYTNA 161
573 CGCCACAGACGCCGAGACAGCTCCGAGGGCAGCGTGAAGCGCTTTCAGACCAAG 632
162 YMSNARCGNGNGARCCNGAIVGNGARMSNTTNGAYGARCRCWMSNWSNWSNWS 221
633 CGGCCCCAGTGTCTCCACAGCTGTGAAGAACACAGCTCTTCTGAAAAGTGA 692
222 NAARMGNNSNTTYTWSNMGNAARTTYMGNGNWSNARMGWSNWSNWSNWSNWS 281
693 AGGAGGGGATGACAGATGACACCTCCGATAGTACACGATGCGCTGACCTGAAGA 752
282 NGNGARMSNGARCMNGATATGTTNGMNSNTTNGMNGVMSGNTTITGTTNA 341
753 GCTTCAGATCCCTTCGAGGAAGCGGAGAACAGAGCCCATGAGAGGCCCTGAAGG 812
342 BAAVGNATGWSNTTMCNCAATYNAAYABARABARAGNGGAGAAAGNACWMSNA 401
813 GATCCAGATCGCTCGGAGAAAGCGCGGAGAGAGGCTCCGCGGAATCTGTGGCTC 872
402 RYTNCNARMSNTTWSNWSNWSNWSNWSNWSNWSNWSNWSNWSNWSNWSNWSNWS 461

873 CGAGGCCAGTACACTGTGAGAGGCGCTCTGCTCCAGTAAAGCAGAGCCCGAGACATCA 932
482 YGARGARCGNGNACNGARGCGNGTNCNMMNGNAAVGGANGCNGANCNCAYMS 521
933 GGGGTTGTGTCCAGCGCTGGGAAGATGACAGAGA 968
522 NCCNAYCCNTTYTNGAYGARGARGCNTTYGANGA 557

RESULT 12

US-09-669-751-79
Sequence 79, Application US/09669751
Patent No. 6551575
GENERAL INFORMATION:
APPLICANT: Greenspan, Ralph J.
TITLE OF INVENTION: Methods for Identifying Compounds for
TITLE OF INVENTION: Motion Sickness, Vertigo and Other Disorders Related to
TITLE OF INVENTION: Balance and the Perception of Gravity
FILE REFERENCE: P-NI 3864
CURRENT APPLICATION NUMBER: US/09/669,751
CURRENT FILING DATE: 2000-09-26
PRIORITY FILING DATE: 1999-12-02
NUMBER OF SEQ ID NOS: 261
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 79
LENGTH: 669
TYPE: DNA
ORGANISM: Drosophila
US-09-669-751-79

Query Match 1.6%; Score 41; DB 4; Length 669;
Best Local Similarity 46.1%; Pred. No. 0.082;
Matches 137; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

742 ACCTGAAAAGCTTCAGATGCTTCGACAGAAAGCGGAAACAGAGCCCACTGAGAG 801
201 ACCTGTGACAGAAATGTAATGCTGCTCCGTCACACGTTTGTGTTCTTCATGATGG 260
802 CCCGTAAGGATGATCAGATGCTGCGGAGAAAGCGCGGAGAGAGGTCGCCGAG 861
261 GGGGATCGAATTTTCTGGAACCATATGAAAGAGAGGCCCAAGAGCGCGCGAGAG 320
862 ACTGTGCTCCGAGGCGAGTACACTGTGAGAGGGGCTCTCCAGTAAAGCAGAGCCC 921
321 GCGCGAATTCGCCAGAGAGAGAGAGAGAGAGGCGCAAGAGCGGAGAGAGAGAG 380
922 GAGACGATCAGGGGCTGTGTCCTCAGGCTGGGAAAGATGACAGAGAGTAAAGTGGAG 981
381 GAAAAGAGAGAGCGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 440
982 GGAAGGCGGCTCAGAGATCAAGATGAGAGGCTGAGAGCTTGGGCGGACCGAG 1038
441 GGAAGGAAAG 497

RESULT 13

US-09-621-976-17202
Sequence 17202, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTe and Encoded Human Proteins.
FILE REFERENCE: GENSET-054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 17202
LENGTH: 364
TYPE: DNA

ORGANISM: Homo sapiens
US-09-621-976-17202

Query Match 1.6%; Score 40.8; DB 4; Length 364;
Best Local Similarity 13.1%; Pred. No. 0.062;
Matches 36; Conservative 123; Mismatches 115; Indels 0; Gaps 0;

426 GCACGACGTCGAGCTTCCTGCGCGGACGAGGAGGAGCCGACGACGCGGT 485
10 GRARCCGCTCGAGAGTSGMCKSSRYGRSSGCCSGMGSGCCSRMSRCKMSMMSM 69
486 GAGACATTCCTGACCATGCGCGCGCGGAGAGGAGCATGCTGTCTCCCTGGA 545
70 MYRSMKYSTKSTGSCYKGGKMACMTCTMGYAMRYVMAISYCMYSYARVYTCYKRM 129
546 GATTCGTGAGACCCGACGCTCTCCCGCCACAGACGCCGACAGCGCTCCGAGCGAG 605
130 WKCYRTRSRGMCMMGASGMSYRSAGSRYSKSGRGMWKKGSRATSKKGRMM 189
606 CGTGAAGGCGCTCTGAGACGAGCGCGCCCGACGCTCTCCACAGCTGTGAAGA 665
190 MKGSRBRATSRVMSMSMYASIRMSMCSASTRMSASCMYTMMSAGSYASCAMTMS 249
666 ACGACGACCTCTTCTGAAAAGTGAAAAGAGGCG 699
250 KYRCATKMSCTYSWYMRASMKSKYKAMSRKSK 283

RESULT 14

US-09-007-005-17
Sequence 17, Application US/0907005B
Patent No. 6258558
GENERAL INFORMATION:
APPLICANT: Szostak, Jack W.
APPLICANT: Roberts, Richard W.
APPLICANT: Liu, Rihne
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
TITLE OF INVENTION: FUSIONS
FILE REFERENCE: 00786/350003
CURRENT APPLICATION NUMBER: US/09/007.005B
CURRENT FILING DATE: 1998-01-14
EARLIER APPLICATION NUMBER: 60/035.963
EARLIER FILING DATE: 1997-01-27
EARLIER APPLICATION NUMBER: 60/064.491
EARLIER FILING DATE: 1997-11-06
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 289
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Translation template
NAME/KEY: misc feature
LOCATION: (1)...(289)
OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 1.5%; Score 40.2; DB 3; Length 289;
Best Local Similarity 5.6%; Pred. No. 0.079;
Matches 12; Conservative 97; Mismatches 106; Indels 0; Gaps 0;

808 AAAGGATCCAGAGTCGCTGCGGAGAAAGCCCGGAGAGAGGCTCCGCGGACGCTGTG 867
51 RARURGRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNS 110
868 GGCTCCGAGCGAGTGAAGTGTGAGGCGCTCTGCGCAAGTGAAGAGAGCCGAGAAC 927
111 RNRNRNRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNS 170
928 GATCAGGGGCTGTGTCTCCAGGCTGGAAAGTGAAGAGAGAGTGAAGTGAAGGAAAG 987

Db 171 RNRNRNRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNS 230
Qy 988 GCGCTCAGACATCAAGATGAGAGGCTTGAGA 1022
Db 231 RCRGRURARCRURCRURURGRGCRGRUAAAAA 265

RESULT 15

US-09-244-796-17
Sequence 17, Application US/09244796
Patent No. 6281344
GENERAL INFORMATION:
APPLICANT: Szostak, Jack W.
APPLICANT: Roberts, Richard W.
APPLICANT: Liu, Rihne
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
TITLE OF INVENTION: FUSIONS
FILE REFERENCE: 00786/350007
CURRENT APPLICATION NUMBER: US/09/244.796
CURRENT FILING DATE: 1999-02-05
EARLIER APPLICATION NUMBER: 60/035.963
EARLIER FILING DATE: 1997-01-27
EARLIER APPLICATION NUMBER: 60/064.491
EARLIER FILING DATE: 1997-11-06
EARLIER APPLICATION NUMBER: 09/007.005
EARLIER FILING DATE: 1998-01-14
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 289
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Translation template
NAME/KEY: misc feature
LOCATION: (1)...(289)
OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match 1.5%; Score 40.2; DB 3; Length 289;
Best Local Similarity 5.6%; Pred. No. 0.079;
Matches 12; Conservative 97; Mismatches 106; Indels 0; Gaps 0;

808 AAAGGATCCAGAGTCGCTGCGGAGAAAGCCCGGAGAGAGGCTCCGCGGACGCTGTG 867
51 RARURGRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNS 110
868 GGCTCCGAGCGAGTGAAGTGTGAGGCGCTCTGCGCAAGTGAAGAGAGCCGAGAAC 927
111 RNRNRNRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNS 170
928 GATCAGGGGCTGTGTCTCCAGGCTGGAAAGTGAAGAGAGAGTGAAGTGAAGGAAAG 987
171 RNRNRNRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNS 230
Qy 988 GCGCTCAGACATCAAGATGAGAGGCTTGAGA 1022
Db 231 RCRGRURARCRURCRURURGRGCRGRUAAAAA 265

Search completed: April 23, 2004, 18:09:01
Job time: 191.89 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 14:17:01 ; Search time 1241.38 Seconds
(without alignments)
9479.366 Million cell updates/sec

Title: US-09-787-016a-1
Sequence: 1 ctccgtgcgcgcgcac.....actcttaagatcatcatctg 2610

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2907579 seqs, 2254313464 residues
Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing filter 45 summaries

Database: Published Applications NA:*

- 1: /cgn2_6/ptcodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptcodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptcodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptcodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptcodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptcodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptcodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptcodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptcodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptcodata/1/pubpna/US09C_PUBCOMB.seq:*
- 11: /cgn2_6/ptcodata/1/pubpna/US09_NEW_PUB.seq:*
- 12: /cgn2_6/ptcodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptcodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptcodata/1/pubpna/US10C_PUBCOMB.seq:*
- 15: /cgn2_6/ptcodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptcodata/1/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptcodata/1/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptcodata/1/pubpna/US60_NEW_PUBCOMB.seq:*
- 19: /cgn2_6/ptcodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1564.4	59.9	1764	9	US-09-925-301-327 Sequence 327, App
2	645	24.7	645	15	US-10-029-386-25091 Sequence 25091, A
3	435	16.7	478	10	US-09-918-995-19127 Sequence 19127, A
4	152.6	5.8	536	15	US-10-029-386-11381 Sequence 11381, A
5	91.2	3.5	226	16	US-10-131-827-8445 Sequence 8445, A
6	65	2.5	5975	10	US-09-873-367C-112 Sequence 112, App
7	65	2.5	6256	15	US-10-037-270-269 Sequence 269, App
8	65	2.5	6256	16	US-10-117-722-269 Sequence 269, App
9	65	2.5	6948	10	US-09-930-213-315 Sequence 315, App
10	60	2.3	60	10	US-09-908-975-6269 Sequence 6269, App
11	52.8	2.0	469	13	US-10-085-783A-56748 Sequence 56748, A
12	52.8	2.0	469	16	US-10-242-535A-56748 Sequence 56748, A
13	52.8	2.0	483	9	US-09-864-761-2140 Sequence 2140, App
14	51.4	2.0	915	9	US-09-764-864-471 Sequence 471, App

15	51.4	2.0	2320	15	US-10-175-523-142	Sequence 142, App
16	50	1.9	50	16	US-10-131-827-916	Sequence 916, App
17	48	1.8	594	13	US-10-112-426-10	Sequence 10, App
18	48	1.8	594	15	US-10-123-155-10	Sequence 10, App
19	48	1.8	594	15	US-10-146-731-10	Sequence 10, App
20	48	1.8	594	15	US-10-140-472-10	Sequence 10, App
21	48	1.8	594	15	US-10-141-761-10	Sequence 10, App
22	48	1.8	594	15	US-10-142-885-10	Sequence 10, App
23	48	1.8	594	15	US-10-158-790-10	Sequence 10, App
24	48	1.8	594	16	US-10-137-871-10	Sequence 10, App
25	48	1.8	594	16	US-10-140-923-10	Sequence 10, App
26	48	1.8	594	16	US-10-141-756-10	Sequence 10, App
27	48	1.8	594	16	US-10-141-759-10	Sequence 10, App
28	48	1.8	594	16	US-10-140-805-10	Sequence 10, App
29	48	1.8	594	16	US-10-140-864-10	Sequence 10, App
30	44.4	1.7	7386	10	US-09-819-104A-6	Sequence 6, App
31	44.4	1.7	7534	13	US-10-087-192-650	Sequence 650, App
32	44.4	1.7	8544	10	US-09-819-104A-4	Sequence 4, App
33	44.4	1.7	91141	13	US-10-087-192-649	Sequence 649, App
34	42.6	1.6	16442	16	US-10-374-077-208	Sequence 208, App
35	42	1.6	2085	17	US-10-311-623-18	Sequence 18, App
36	42	1.6	2771	13	US-10-112-944-494	Sequence 494, App
37	42	1.6	3195	15	US-10-205-823-143	Sequence 143, App
38	42	1.6	3195	15	US-10-285-976-54	Sequence 54, App
39	42	1.6	3195	16	US-10-295-027-1127	Sequence 1127, App
40	42	1.6	3206	13	US-10-112-944-6	Sequence 6, App
41	42	1.6	1223197	13	US-10-027-632-179264	Sequence 179264, App
42	42	1.6	1223197	16	US-10-027-632-179264	Sequence 179264, App
43	41.8	1.6	636	9	US-09-770-149-640	Sequence 640, App
44	41.8	1.6	659158	9	US-09-771-208-20	Sequence 20, App
45	41.2	1.6	4056	15	US-10-156-761-2317	Sequence 2317, App

ALIGNMENTS

RESULT 1
US-09-925-301-327
Sequence 327, Application US/09925301
Patent No. US2002052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 327
LENGTH: 1764
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1398)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1758)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1759)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1762)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-327
Query Match 59.9%; Score 1564.4; DB 9; Length 1764;
Best Local Similarity 99.6%; Pred. No. 0;

Matches	1610;	Conservative	0;	Mismatches	2;	Indels	5;	Gaps	4;
QY	996	GGACATCAAAAGATGAGAGCCTGAGAC- TTGGGCCGACCGAAGCTTGATGTGAGGTT							1054
Db	1	GGACATCAAAAGATGAGAGCCTGAGAGCCTTGGGGCGACCGAAGCCTGAATGTGAGGTT							60
QY	1055	ACGACCCCAAGCCCTGTATGATGATTTGGCCGACGCTCAAGAAAGATTTATGATTT							1114
Db	61	ACGACCCCAAGCCCTGTATGATGATTTGGCCGACGCTCAAGAAAGATTTATGATTT							120
QY	1115	GCTGTGACCGCTGTGAGAGATGTTTCAATGCGATTTGTGTGGGCAATTTCTGAGGCTCGAG							1174
Db	121	GCTGTGACCGCTGTGAGAGATGTTTCAATGCGATTTGTGTGGGCAATTTCTGAGGCTCGAG							180
QY	1175	GGAGGCTTTTGGAAAGAAATGGGGAAAGCTATATCTGCCAACTGCAACATTTCTGCAAG							1234
Db	181	GGAGGCTTTTGGAAAGAAATGGGGAAAGCTATATCTGCCAACTGCAACATTTCTGCAAG							240
QY	1235	TGCAGGATGAGATCAATTCAGAAACGGAGATCAGCAGGAAAGCTAAATGAGACCTGGAG							1294
Db	241	TGCAGGATGAGATCAATTCAGAAACGGAGATCAGCAGGAAAGCTAAATGAGACCTGGAG							300
QY	1295	ATGCTGATGGACCCGATTTGTAACAATATAGAAACAATAGAGAGAAATCTAGCGAAGAC							1354
Db	301	ATGCTGATGGACCCGATTTGTAACAATATAGAAACAATAGAGAGAAATCTAGCGAAGAC							360
QY	1355	AAGGATTAAGGATTAAGATTTGAGAAAGCTGCAATTCGAATGAGCAAGAAAGAACTCAAGA							1414
Db	361	AAGGATTAAGGATTAAGATTTGAGAAAGCTGCAATTCGAATGAGCAAGAAAGAACTCAAGA							420
QY	1415	TCTTCAGACCTGTGATAGAGGCGCTGGTGCCTCAAAATGTATTTGGCCCGGGCTGTGC							1474
Db	421	TCTTCAGACCTGTGATAGAGGCGCTGGTGCCTCAAAATGTATTTGGCCCGGGCTGTGC							480
QY	1475	ACGTGGCGCAGCCCGACTCGGTGTACTGCAATGATGACTGTATCTCAACACGCGCAG							1534
Db	481	ACGTGGCGCA- CCGACTCGGTGTACTGCAATGATGACTGTATCTCTCAACACGCGCAG							539
QY	1535	CGACATTAAGATTTCTAAGCTCAAGTAAAGAAACAGAACCTTAAGAAAGAAAGATGA							1594
Db	540	CGACATTAAGATTTCTAAGCTCAAGTAAAGAAACAGAACCTTAAGAAAGAAAGATGA							599
QY	1595	AGATGAAGCCAGAAAGCCAGTCTTCGAAATGCGGTCTCAGGACAGTATTAATCT							1654
Db	600	AGATGAAGCCAGAAAGCCAGTCTTCGAAATGCGGTCTCAGGACAGTATTAATCT							659
QY	1655	CTTCTGTGCACAAGAGACAGCTCCAGAAAAAAGAGACCAAGTGAAGAGGACGTGG							1714
Db	660	CTTCTGTGCACAAGAGACAGCTCCAGAAAAAAGAGACCAAGTGAAGAGGACGTGG							719
QY	1715	TGGTCCCTGCGCGAGTGAAGCACTCGGAAAGAAAGCAAGCTTTGTAGACAGACGCCGT							1774
Db	720	TGGTCCCTGCGCGAGTGAAGCACTCGGAAAGAAAGCAAGCTTTGTAGACAGACGCCGT							779
QY	1775	CGTGGGCGAGGATCAACAATTACATGCAATGAAGCCAGAAAGACTGCTGCTCCCTGCG							1834
Db	780	CGTGGGCGAGGATCAACAATTACATGCAATGAAGCCAGAAAGACTGCTGCTCCCTGCG							839
QY	1835	CGTCACTGTATTAATGATGATGATACCTAGGGGTTGGCTCTCTGAGACCCCTCCGTT							1894
Db	840	CGTCACTGTATTAATGATGATGATACCTAGGGGTTGGCTCTCTGAGACCCCTCCGTT							899
QY	1895	CTTTCTGATAGCCATCCCTGCGGCTGTCCAGGACTGGGAGTTGCACTTTGTGTTAAG							1954
Db	900	CTTTCTGATAGCCATCCCTGCGGCTGTCCAGGACTGGGAGTTGCACTTTGTGTTAAG							959
QY	1955	CTGATCAAGACACCGGCTGACCATCAGCGGAAAGAGAGCCCATGTCAGAGATGCTTC							2014
Db	960	CTGATCAAGACACCGGCTGACCATCAGCGGAAAGAGAGCCCATGTCAGAGATGCTTC							1019
QY	2015	CTGCTGCGCTGTGTCCATCCCTAGTCTGTCAAGACTTCTGTCACTGTTTCAAAAGCTG							2074
Db	1020	CTGCTGCGCTGTGTCCATCCCTAGTCTGTCAAGACTTCTGTCACTGTTTCAAAAGCTG							1079

QY	2075	TAACCTCAGTGTGAAGCTTCACTTAATGATGATTTCTTAATCTGTTTCACTCT							2134
Db	1080	TAACCTCAGTGTGAAGCTTCACTTAATGATGATTTCTTAATCTGTTTCACTCT							1139
QY	2135	CAGGCTGTGATGATTTGATTTCTCTTCAATCCAGTCTGATGATGATGACACACTGCC							2194
Db	1140	CAGGCTGTGATGATTTGATTTCTCTTCAATCCAGTCTGATGATGATGACACACTGCC							1199
QY	2195	CGGACGCAATCCACCCCTGTCTGCAATGATGTTTGTGACAAACAGGCTGTATAG							2254
Db	1200	CGGACGCAATCCACCCCTGTCTGCAATGATGTTTGTGACAAACAGGCTGTATAG							1259
QY	2255	CTTCAGTTTTCACATGTTGCAAGGCGGACAGACATGAAGATCACTTTTATGTT							2314
Db	1260	CTTCAGTTTTCACATGTTGCAAGGCGGACAGACATGAAGATCACTTTTATGTT							1319
QY	2315	GTGGGAATCTTTCAGATGATGATGATGATTTTCAAGGTGATCAATTTATTTGACT							2374
Db	1320	GTGGGAATCTTTCAGATGATGATGATGATTTTCAAGGTGATCAATTTATTTGACT							1379
QY	2375	GGGAGATAGGGGATTTTTTTTTTTTTCATGTCGATTCACACCTTACACCCACATG							2434
Db	1380	GGGAGATAGGGGATTTTTTTTTTTTTCATGTCGATTCACACCTTACACCCACATG							1437
QY	2435	AACCATTCGAATCTTGAAAGG- CACACACTGCTGATGAGGCGCCACGGTAAGT							2493
Db	1438	AACCATTCGAATCTTGAAAGG- CACACACTGCTGATGAGGCGCCACGGTAAGT							1497
QY	2494	TCACACTAGAACCTGTCTGACCGCAGAGACGCTGCTTGATGAGCTGTATTTCAATG							2553
Db	1498	TCACACTAGAACCTGTCTGACCGCAGAGACGCTGCTTGATGAGCTGTATTTCAATG							1557
QY	2554	TGACTGCTTTCTTGGCTCTGCTCTTGAATTTAAGCTCTTAAGATCATCTCTG							2610
Db	1558	TGACTGCTTTCTTGGCTCTGCTCTTGAATTTAAGCTCTTAAGATCATCTCTG							1614

RESULT 2

US-10-029-386-25091/c
 ; Sequence 25091, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI
 ; FILE REFERENCE: AEMICA-X-2
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 25091
 ; LENGTH: 645
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURES:
 ; OTHER INFORMATION: MAP TO AL117379.14
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
 ; OTHER INFORMATION: SWISSPROT HIT: P23583, EVALU2 2.50e-01
 ; OTHER INFORMATION: EST HUMAN HIT: AL580669.1, EVALU8 0.00e+00
 ; OTHER INFORMATION: NT HIT: g16193859, EVALU6 0.00e+00
 US-10-029-386-25091

Query Match

24.7%; Score 645; DB 15; Length 645;
 Best Local Similarity 100.0%; Pred. No. 1.5e-196;

Matches 645; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	263	GTATGACGACAAAGGCGACCCGAGCAATGAGAGGACCTTAAGGCTCAAAACCAACA							322
Db	645	GTATGACGACAAAGGCGACCCGAGCAATGAGAGGACCTTAAGGCTCAAAACCAACA							586

323 GCAAGAGTTCAGGAAAAATGCGGTTTTCGAGAGACATATCGCCAGGAGAGGCG 382
585 GCAAGAGTTCAGGAAAAATGCGGTTTTCGAGAGACATATCGCCAGGAGAGGCG 526
383 CAGGGAGCGGAGGCTGAGACCTGAGAGCCGCCACAGAGAGGCTGGGCTGT 442
525 CAGGGAGCGGAGGCTGAGACCTGAGAGCCGCCACAGAGAGGCTGGGCTGT 466
443 CCGTGGCGGCGAGTGGAGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 502
465 CCGTGGCGGCGAGTGGAGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 406
503 TTGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 562
405 TTGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 346
563 CGTCTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 622
345 CGTCTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 286
623 AGAGCAGAGGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 682
285 AGAGCAGAGGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 226
683 AAAAGGTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 742
225 AAAAGGTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 166
743 CTTGAAAGAGCTTCAGAGTCCCTGTCGAGAGAGGAGGAGGAGGAGGAGGAGG 802
165 CTTGAAAGAGCTTCAGAGTCCCTGTCGAGAGAGGAGGAGGAGGAGGAGGAGG 106
803 CCGTGAAGAGGATCCAGAGTCCCTGTCGAGAGAGGAGGAGGAGGAGGAGGAGG 862
105 CCGTGAAGAGGATCCAGAGTCCCTGTCGAGAGAGGAGGAGGAGGAGGAGGAGG 46
863 CTTGAGGCTCCGAGGCGGAGTGAACCTGAGAGGAGGAGGAGGAGGAGGAGGAGG 907
45 CTTGAGGCTCCGAGGCGGAGTGAACCTGAGAGGAGGAGGAGGAGGAGGAGGAGG 1

RESULT 3
US-09-918-995-19127
; Sequence 19127, Application US/09918995
; Publication No. US2003007623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 19127
LENGTH: 478
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1) - (478)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-19127

Query Match 16.7%; Score 435; DB 10; Length 478;
Best Local Similarity 100.0%; Pred. No. 6.8e-129;
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

905 CCAAGTAAAGCAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 964
44 CCAAGTAAAGCAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 103

965 GAGAGAGTAAAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1024
104 GAGAGAGTAAAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 163
1025 TGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1084
164 TGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 223
1085 GCCAGCTCACAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1144
224 GCCAGCTCACAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 283
1145 GCGATGTTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1204
284 GCGATGTTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 343
1205 ATATCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1264
344 ATATCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 403
1265 ATCAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1324
404 ATCAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 463
1325 GAACATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1339
464 GAACATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 478

RESULT 4
US-10-029-386-11381/c
; Sequence 11381, Application US/10029386
; Publication No. US2003019470A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI
TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
FILE REFERENCE: ABOVICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 11381
LENGTH: 536
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL117379.14
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
OTHER INFORMATION: SWISSPROT HIT: P80398, EVALU8 3.80e+00
OTHER INFORMATION: NT HIT: gi14769835, EVALU8 1.00e-78
OTHER INFORMATION: EST_HUMAN HIT: BF205331.1, EVALU8 2.00e-78
US-10-029-386-11381

Query Match 5.8%; Score 152.6; DB 15; Length 536;
Best Local Similarity 94.6%; Pred. No. 9.5e-38;
Matches 158; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

244 GCTTACGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 303
167 GCTTACGTTTACGCTAGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 108
304 AAGGCGATCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 363
107 AAGGCGATCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 48
364 ATGCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 410
47 ATGCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 112
; LENGTH: 5975
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-873-367C-112

Query Match
Best Local Similarity 62.7%; Score 65; DB 10; Length 5975;
Matches 101; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1060 CCCAAGCCCTGTATATTCATTGCGCCAGCCTCAACAACAAGTTATGATTTGCTGT 1119
    |||||
DB 1200 CCCAGACAGAGATGTGGGTTTTCACAAAACCAACAGGATTAAGTTGCTGT 1259
    |||||

QY 1120 GACCGCTGTGAAGATGCTTTCATGGCGATTGTGTGGCATTTCTGAAGCTCGAAGAG 1179
    |||||
DB 1260 GGGAGATGTATATACATGCTTTCATGGGATGTGTGTGGTAAAGTCTTCTCAAGCAG 1319
    |||||

QY 1180 CTTTGGAAAGGAATGGGGAAGACTATATCTGCGCAACTG 1220
    |||||
DB 1320 CAGATGGCGAGAGACAAAGAAATATGTCTGTATTAATG 1360
    |||||

RESULT 7
; Sequence 269, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyun
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunhui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784C1P2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pc FL_genes Version 1.0
; SEQ ID NO 269
; LENGTH: 6256
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (257) .. (5443)
US-10-037-270-269

Query Match
Best Local Similarity 62.7%; Score 65; DB 15; Length 6256;
Matches 101; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1060 CCCAAGCCCTGTATATTCATTGCGCCAGCCTCAACAACAAGTTATGATTTGCTGT 1119
    |||||
DB 1469 CCCAAGACAGATGTGGGTTTTCACAAAACCAACATGGCAACAGTTATGCTGT 1528
    |||||

```

QY	Db	QY	Db
1120	GCACGCTGGAAGAAATGGTTTATCGGATTTGTGGCAATTCCTGAGAGCGTCGAGGGAG	1179	
1529	GGGAGATGTATATACCTGCTTCATCGCATTTGTGTGGCTTAATCTTTCCTCAAGCAG	1588	
1180	CTTTTGGAAAGAAATGGGGAAGACTATATCTGCCCAACTG	1220	
1589	CAGATGGCGAGGACAGCAAGATATGTCTGTCTTAATG	1629	

RESULT 8
US-10-117-722-269
; Sequence 269, Application US/10117722
Publ. Ref. No. US200201874A1

Query Match	2.5%	Score 65	DB 16	length 6256
Best local Similarity	62.7%	Pred. No.	9 4e-09	
Matches 101, Conservative	0	Mismatches	60	Indels 0
				Gaps 0

US-00-930-213-315
Sequence 315, Application US/09930213
Publication No. US20030170625A1
GENERAL INFORMATION:
APPLICANT: ROSENTHAL, ANDRE
HINZMANN, BERND
APPLICANT: SCHAFFER, REINHARD
APPLICANT: ZOBBER, JOHANNES
APPLICANT: TCHÉ-NITSE, OLEG
APPLICANT: GRIPS, MARTIN
APPLICANT: HELINGELG, MARTIN
APPLICANT: SCHMITZ, ANNE-CHANTAL
APPLICANT: SERS, CHRISTINE

```

1  TITLE OF INVENTION: DETECTION OF DIFFERENTIAL GENE EXPRESSIONS
2
3  FILE REFERENCE: ADURE-14
4
5  CURRENT APPLICATION NUMBER: US/09/930.213
6
7  CURRENT FILING DATE: 2001-01-31
8
9  PRIOR APPLICATION NUMBER: DB 10004102.7
10
11 PRIOR FILING DATE: 2000-01-31
12
13 NUMBER OF SEQ ID NOS: 885
14
15 SOFTWARE: PatentIn Ver. 2.1
16
17 SEQ ID NO 315
18
19 LENGTH: 6948
20
21 TYPE: DNA
22
23 ORGANISM: Homo sapiens
24
25 US-09-930-213-315

```

Qy	1180	CTTTGGAMAGGATGGGAGACTATATCTGCCCAACTG	1220
Db	2293	CAGATGGCGAGAGACAAAGAATATGTCTGTGTMAATG	2333

```

; TYPE: DNA
; ORGANISM: Homo sapiens
DS-09-908-975-6269

```

RESULT 11
US-10-085-783A-56748
; Sequence 56748, Application US/10085783A
; Publication NO. US2004003784A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.

APPLICANT: Liew, C.C.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2002
CURRENT APPLICATION NUMBER: US/10/085,783A
CURRENT FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 56748
LENGTH: 469
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (5)..(5)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (30)..(30)
OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-56748

Query Match 2.0%; Score 52.8; DB 13; Length 469;
Best Local Similarity 65.0%; Pred. No. 1.4e-05;
Matches 78; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1101 CAGGTTATGATTTGCTGTGACCGCTGTGAAGATGTTTCATGCGGATTTGGGCAT 1160
DB 116 CAGGTTATGATTTGCTGTGACCGCTGTGAAGATGTTTCATGCGGATTTGGGCAT 1160
QY 1161 TTCTGAGCTCGAGGAGCTTTTGAAGAATGCGGATTTATCTGCGCAACTG 1220
DB 176 AAGCTTTCTCAAGCAGCAGATGCGGAGGAGCAAGAATATGCTGTGTAATG 235

RESULT 12
US-10-242-535A-56748
Sequence 56748, Application US/10242535A
Publication No. US20040013663A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
APPLICANT: Liew, C.C.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2005
CURRENT APPLICATION NUMBER: US/10/242,535A
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 10/085,783
PRIOR FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 56748
LENGTH: 469
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (5)..(5)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (30)..(30)
OTHER INFORMATION: n is a, c, g, or t

US-10-242-535A-56748

Query Match 2.0%; Score 52.8; DB 16; Length 469;
Best Local Similarity 65.0%; Pred. No. 1.4e-05;
Matches 78; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1101 CAGGTTATGATTTGCTGTGACCGCTGTGAAGATGTTTCATGCGGATTTGGGCAT 1160
DB 116 CAGGTTATGATTTGCTGTGACCGCTGTGAAGATGTTTCATGCGGATTTGGGCAT 1160
QY 1161 TTCTGAGCTCGAGGAGCTTTTGAAGAATGCGGATTTATCTGCGCAACTG 1220
DB 176 AAGCTTTCTCAAGCAGCAGATGCGGAGGAGCAAGAATATGCTGTGTAATG 235

RESULT 13
US-09-864-761-2140
Sequence 2140, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 2140
LENGTH: 483
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL050329..9

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 07:20:53 ; Search time 6433.74 Seconds
(without alignment)

12114.300 Million cell updates/sec

Title: US-09-787-016a-1

Sequence: 1 ctcgtgagccgcgcgcacccac.....actcttaagatcatatccctg 2610

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthm: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hci: *
9: gb_est1: *
10: gb_est2: *
11: gb_hci: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: em_gss_hum: *
18: em_gss_hiv: *
19: em_gss_p1n: *
20: em_gss_vrt: *
21: em_gss_fun: *
22: em_gss_mam: *
23: em_gss_mus: *
24: em_gss_pro: *
25: em_gss_rtd: *
26: em_gss_pbg: *
27: em_gss_vrl: *
28: gb_gss1: *
29: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2539	97.3	2752	11	BC014615
2	1025.2	39.3	4619	9	AK042474
3	1007.2	38.6	1086	9	AL578222
4	1002.6	38.4	4651	11	BC060249

5	957.2	36.7	1093	9	AL555810
6	910.6	34.9	1033	9	AL580669
7	905.6	34.7	1201	9	AL576399
8	900.2	34.5	1010	13	BX380892
9	868	33.3	1201	9	AL514138
10	863.6	33.1	1201	9	AL553387
11	845.2	32.4	1018	9	AL550675
12	838.2	32.1	1050	13	BX404270
13	830.4	31.8	850	14	CD364822
14	817.8	31.3	1614	11	AK014422
15	813.4	31.2	1119	9	AL550431
16	808.6	31.0	1640	11	BC029110
17	798.6	30.6	898	13	B0646878
18	792.8	30.4	939	9	AL571596
19	786.6	30.1	891	13	BX350413
20	786.4	30.1	1040	9	AL580668
21	786.2	30.1	925	9	AL546121
22	765.8	29.3	1201	13	BX442517
23	753.4	28.9	781	13	B0921479
24	740.8	28.4	1061	9	AL574291
25	738.8	28.3	842	13	BQ437271
26	733.2	28.1	866	13	BQ212019
27	722.4	27.7	1201	9	AL514137
28	717	27.5	1045	9	AL550704
29	708.4	27.1	740	13	B0689908
30	693.8	26.6	779	12	B0825278
31	673.2	25.8	842	12	B0757160
32	670.6	25.7	839	12	B0828423
33	668.6	25.6	1097	12	B0290255
34	668.2	25.6	900	13	BX411695
35	648.6	24.9	760	10	B0298017
36	645	24.7	740	12	B1913500
37	634.4	24.3	963	13	B0526593
38	629.6	24.1	772	10	B0299127
39	626.8	24.0	692	10	B0393471
40	625	23.9	635	13	B0639998
41	615.6	23.6	653	10	AM373892
42	615.6	23.6	653	10	AM373829
43	614.4	23.5	923	10	B0308603
44	612.4	23.5	1036	13	B0881380
45	608.8	23.3	868	10	B0306721

ALIGNMENTS

RESULT 1
BC014615
LOCUS
DEFINITION
ACCESSION
BC014615
VERSION
BC014615.2
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 2752)
Strausberg,R.L., Feingold R.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,P.S., Wagner,L., Shennan,C.M., Schler,G.D.,
Altschul,S.F., Zeeberg,B., Buecaw,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,Y., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.P., Casavant,T.L.,
Schetz,T.R., Brownstein,M.J., Utsid,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Muliyil,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Huiyik,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,B.J., Lu,X., Gibbs,R.A.,
Fahy,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITL
JOURNAL

12477932
2 (bases 1 to 2752)
Strauberg, R.
Direct Submission
Submitted (24-SEP-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Aug 19, 2003 this sequence version replaced gi:15779097.
Contact: MGC help desk
Email: cgapdb-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahy, Erin Helton, Mark Ketteman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

REMARK
COMMENT

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>
Series: IRML Plate: 11 Row: j Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 18375616
This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
1..2752
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3532298"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_1ib="NIH MGC 17"
/lab_host="DH10B-R"
/note="vector: pOTB7"

ORIGIN

Query Match 97.3%; Score 2539; DB 11; Length 2752;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2573; Conservative 0; Mismatches 0; Indels 4; Gaps 3;

36 GAAATGGCTGCGAGACCTTAGAGGCTGCGAGGCTTACTCCAGGGAACAGCCTTAGAT 95
1 GAAATGGCTGCGAGACCTTAGAGGCTGCGAGGCTTACTCCAGGGAACAGCCTTAGAT 60
96 AATCTGAGTTGTGAAATAACGAAGCTTGTACTGTGAACAGTGGCTGCAACATGTT 155
61 AATCTGAGTTGTGAAATAACGAAGCTTGTACTGTGAACAGTGGCTGCAACATGTT 120
156 GTTGGAGCTGGCTGTCTGCTTGAACCAAGAGGTTGCTGCGAGGTTTGGTTGT 215
121 GTTGGAGCTGGCTGTCTGCTTGAACCAAGAGGTTGCTGCGAGGTTTGGTTGT 180
216 ATTAGAGATTTCAGGAAAGTGTCCAGCTTTCAGTGTGGAGCAGGTATGAGCAAA 275
181 ATTAGAGATTTCAGGAAAGTGTCCAGCTTTCAGTGTGGAGCAGGTATGAGCAAA 240
276 AGCCGACCCGAGCAATGAGAGGCACTTAAGGCATCAAAACCAACAGAAAGATTTCAG 335
241 AGCCGACCCGAGCAATGAGAGGCACTTAAGGCATCAAAACCAACAGAAAGATTTCAG 300
336 GAAACATATGGGTTTTCGAAAGACCACTATGCGCAAGCAGAGGCGCAGGGAGCCGGA 395
301 GAAACATATGGGTTTTCGAAAGACCACTATGCGCAAGCAGAGGCGCAGGGAGCCGGA 360

396 GGCTTACCCACTGAGAGCCGCCAACCACAGAGCAGGCTGGCTGTCCCTGGCGGCGAG 455
361 GGCTTACCCACTGAGAGCCGCCAACCACAGAGCAGGCTGGCTGTCCCTGGCGGCGAG 420
456 TGGAGAGCAGCCCAAGCCCACTAGACGCGTGGAGCAGTTCTTCAACATTTGCGCGCGCG 515
421 TGGAGAGCAGCCCAAGCCCACTAGACGCGTGGAGCAGTTCTTCAACATTTGCGCGCGCG 480
516 CCGCAGAGAGAGCAGTCTGTCTCTCTTGGAGAGTTCTGTGAGCCCAAGTCTCTGCGCGCG 575
481 CCGCAGAGAGAGCAGTCTGTCTCTCTTGGAGAGTTCTGTGAGCCCAAGTCTCTGCGCGCG 540
576 CACACAGCCCGAGAGAGCCTCCGAGAGCAGCGTGGAGAGCGTTCTGAGACAGAGAGCG 635
541 CACACAGCCCGAGAGAGCCTCCGAGAGCAGCGTGGAGAGCGTTCTGAGACAGAGAGCG 600
636 CCCCAGTCTGTCTTCCAGAGCTGTGAGAGAGCAGCAGCCTCTTGTGAAAGGTGAAAG 695
601 CCCCAGTCTGTCTTCCAGAGCTGTGAGAGAGCAGCAGCCTCTTGTGAAAGGTGAAAG 660
696 AGGGAGATGACCAAGATGACACCTCGGATATGACAGCGATGACCTTGAAGAGCT 755
661 AGGGAGATGACCAAGATGACACCTCGGATATGACAGCGATGACCTTGAAGAGCT 720
756 TGAAGATGCTCTTCCGAGAGACCGGGAACAGAGGCCCAATTGAGAGCGCCCTGAAAGGAGT 815
721 TGAAGATGCTCTTCCGAGAGACCGGGAACAGAGGCCCAATTGAGAGCGCCCTGAAAGGAGT 780
816 CAGAGTGCCTGCGGAGAGAGCGCGGAGAGAGGCTCCGCGAGACTGTGGGCTCCGA 875
781 CAGAGTGCCTGCGGAGAGAGCGCGGAGAGAGGCTCCGCGAGACTGTGGGCTCCGA 840
876 GGCAGATGACACTGTGAGAGGCGCTCTCTGCGAGTAAAGCAGAGCCTGAGAGATCAAGG 935
841 GGCAGATGACACTGTGAGAGGCGCTCTCTGCGAGTAAAGCAGAGCCTGAGAGATCAAGG 900
936 GGTGTGTCTCCAGCGCTGGGAGAAAGATGACAGAGAGTAAAGTTGAGAGGAAAGCGGCTCA 995
901 GGTGTGTCTCCAGCGCTGGGAGAAAGATGACAGAGAGTAAAGTTGAGAGGAAAGCGGCTCA 960
996 GGAACATCAAGATAGAGAGCCTTGAAGCTTGGGCGCAGCCGAGAGCTGAAATGTGAGGTTA 1055
961 GGAACATCAAGATAGAGAGCCTTGAAGCTTGGGCGCAGCCGAGAGCTGAAATGTGAGGTTA 1020
1056 CGAACCCCAACCGCCTGTATTTGCAATTTGCCGCGAGCTTCAACACAGAGTTATGATTTG 1115
1021 CGAACCCCAACCGCCTGTATTTGCAATTTGCCGCGAGCTTCAACACAGAGTTATGATTTG 1080
1116 CTGTGACCGCTGTGAAAGATGCTTTCAATGAGCCATTTGTGTGGCATTTCTGAGGCTGAGG 1175
1081 CTGTGACCGCTGTGAAAGATGCTTTCAATGAGCCATTTGTGTGGCATTTCTGAGGCTGAGG 1140
1176 GAGGCTTTTGAAGAAAGATAGGGAAGACTATATCTGCCAAACTGACCAATTTCTGCAAGT 1235
1141 GAGGCTTTTGAAGAAAGATAGGGAAGACTATATCTGCCAAACTGACCAATTTCTGCAAGT 1200
1236 GCAGATGAGACTCATTCAGAAACGCGCAGATCAGCAGAACTTAAATGAGACTTGGAGA 1295
1201 GCAGATGAGACTCATTCAGAAACGCGCAGATCAGCAGAACTTAAATGAGACTTGGAGA 1260
1296 TGTGTATGCAACCGATTTGTAAAGATATAGAAACAAATAGAGCAGAGCTTAGAGAGCA 1355
1261 TGTGTATGCAACCGATTTGTAAAGATATAGAAACAAATAGAGCAGAGCTTAGAGAGCA 1320
1356 AGGAGTAAAGGGTATGAGAAAGCTCAATATCCAGTGGCAGAGAACTCAAGAT 1415
1321 AGGAGTAAAGGGTATGAGAAAGCTCAATATCCAGTGGCAGAGAACTCAAGAT 1380
1416 CTTCAGGCTGTGATGAGAGCGCTGTGTCTTCAAAATGATATTTGGCCCGGGGTCTGTCA 1475
1381 CTTCAGGCTGTGATGAGAGCGCTGTGTCTTCAAAATGATATTTGGCCCGGGGTCTGTCA 1440

QY	1476	CGTGGCGGACGCCGACCTCGGTGTACTGCACTAATGATCTGTAATCTCTGAAACAACCCCGAGC	1533
DB	1441	CGTGGCGCACCCGACCTCGGTGTATCTGCAGTAATGACTGTATCTCTCAACACGCGCAGC	1500
QY	1536	GACAAATGAACTTTCTAAGCTCAGGTAAAGAACAGAACCCAAAGCCTTAAAGAAAAGTAA	1595
DB	1501	GACATATGAGTTTCTAAGCTCAGGTAAAGAACAGAACCCAAAGCCTTAAAGAAAAGTAA	1560
QY	1596	GATGAAGCCAGAGAGAGCCAGTCTTCCGAAATGCGGTGCTCAAGCAGTATTTAAATCTC	1655
DB	1561	GATGAAGCCAGAGAGAGCCAGTCTTCCGAAATGCGGTGCTCAAGCAGTATTTAAATCTC	1620
QY	1656	TTCTGTGTCAAGAGACCACTCCAG--AAAAAAAAGACCAACGTATAAAGGCACTGG	1714
DB	1621	TTCTGTGTCAAGAGACCACTCCAGAAAAAAAAGACCAACGTATAAAGGCACTGG	1680
QY	1715	TGTCCTCGCGCGGAGTGAAGCACTCGGGAAGAAAGACGTTGTGAGAGCAGACGCGT	1774
DB	1681	TGTCCTCGCGCGGAGTGAAGCACTCGGGAAGAAAGACGTTGTGAGAGCAGACGCGT	1740
QY	1775	CGTGGCGGAGGATCACAATTACAATGCAATGAAGCCAGAAAAGACTGCTGCTCCCTGC	1834
DB	1741	CGTGGCGGAGGATCACAATTACAATGCAATGAAGCCAGAAAAGACTGCTGCTCCCTGC	1800
QY	1835	CGTCACTGTTGTATTAATGTAATGTAATCACTTAGGGGTGGGCTCTGGAACCCCTCCGTT	1894
DB	1801	CGTCACTGTTGTATTAATGTAATGTAATCACTTAGGGGTGGGCTCTGGAACCCCTCCGTT	1860
QY	1895	CTTTCTGGATGAGCAATCCCTGAGGCTGTCCAGGACTGGAGTTGCACTTGTGTAA	1954
DB	1861	CTTTCTGGATGAGCAATCCCTGAGGCTGTCCAGGACTGGAGTTGCACTTGTGTAA	1920
QY	1955	CTGATCAAGACACCGGCTGACCACTACGCGGGAAGCAGAGCCCAATGTCCAGATGCTC	2014
DB	1921	CTGATCAAGACACCGGCTGACCACTACGCGGGAAGCAGAGCCCAATGTCCAGATGCTC	1980
QY	2015	CTGCTGCCCTGTGTTCATCCCTAGTCTGTCAAGACTTCTGTCACTGTTTCCAAAGCTG	2074
DB	1981	CTGCTGCCCTGTGTTCATCCCTAGTCTGTCAAGACTTCTGTCACTGTTTCCAAAGCTG	2040
QY	2075	TAAACCTACGTGTGAAGGTCAACCTTAATGATGTAATCTTAAATCTGTTTCACTCT	2134
DB	2041	TAAACCTACGTGTGAAGGTCAACCTTAATGATGTAATCTTAAATCTGTTTCACTCT	2100
QY	2135	CAGGCTCGTAAAGTATTTGTAATCTCTTCACTCCAGTGTGATTCATAGCACACTGCC	2194
DB	2101	CAGGCTCGTAAAGTATTTGTAATCTCTTCACTCCAGTGTGATTCATAGCACACTGCC	2160
QY	2195	CGGACGCGCAATCCACCCCTGTCTGCACATGATGTTTCTGACAACACGCGTGTATAG	225
DB	2161	CGGACGCGCAATCCACCCCTGTCTGCACATGATGTTTCTGACAACACGCGTGTATAG	2220
QY	2255	CTTCACTTTTCCAAATGTCCAGGCGCAGCAGTAAAGATCACTCTTTTAAATGTT	2314
DB	2221	CTTCACTTTTCCAAATGTCCAGGCGCAGCAGTAAAGATCACTCTTTTAAATGTT	2280
QY	2315	GTTGGAAATCTTTCGAAAGTATGTTGATCACTGATTTTCAAGTGTACATTTATTTTGA	2374
DB	2281	GTTGGAAATCTTTCGAAAGTATGTTGATCACTGATTTTCAAGTGTACATTTATTTTGA	2340
QY	2375	GCGCAGATAGGGGATTTTTTTTTTTTCCATGTCCGATTTCAACGCTACACCCCAATG	2433
DB	2341	GCGCAGATAGGGG--TTTTTTTTTTCCATGTCCGATTTCAACGCTACACCCCAATG	2398
QY	2435	AAACAATTCGAATTCCTGAAAG--CAACAATCCTGTCTTAATGGGCCCCAGGTAATGAGT	2499
DB	2399	AAACAATTCGAATTCCTGAAAG--CAACAATCCTGTCTTAATGGGCCCCAGGTAATGAGT	2465
QY	2494	TCACACTTAGAACACTGTCTTGACCGCAGAGCGAGTCCCTTGGAATTGGTATTTCAATG	2553
DB	2459	TCACACTTAGAACACTGTCTTGACCGCAGAGCGAGTCCCTTGGAATTGGTATTTCAATG	2518
QY	2554	TGACCTGGCTTTCTTGCCCTCGTCTCTTAATGTTAAGCTCTTAAGATCAATCTCTG	2610

[illegible]

COMMENT

Kanagawa 230-0045, Japan (E-mail: genome-res@isc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
 Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
 Trust/MRC building Addenbrookes Hospital Cambridge) whose
 assistance we gratefully acknowledge.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Location/Qualifiers

FEATURES

source

CDS

1..4619
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:A630095A07"
 /db_xref="MGI:2407519"
 /db_xref="taxon:10090"
 /clone="A630095A07"
 /tissue_type="thymus"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="3 days neonate"
 299..2143
 /note="unnamed protein product; death
 inducer-oblierator-1 (MGI:MGI:1344352, GI:A238332,
 evidence: BLASTN, 99%, match=2599)
 putative"
 /codon_start=1
 /protein_id="BAC31270.1"
 /db_xref="GI:26335139"
 /translation="MDDKGLSNBAPKAIKPTSKFRRTMGFRRTIAKRGAGDTE
 VPSRQPOQHNLSTLRSGRQKRTKRYEPTTTPRRKKNVPLVLSGSRSTV
 TDVEATSGSVSSSIRSGPDSLGKHPASSEKAKGSEBBDSDSDSLTKE
 LQNLRRKROBVERSLSQNLKRRRBDASVQJGSAPDPLKQBEA
 SGPVQSRETDIENLBKATQGNBENRBRKPRCEYDPAALCICROPNN
 RMIICDRBWFHDCVIGISARGLKRNGBDYICPCTLIQVDEYNGSATDQD
 SCRSVAGDGTCTSIGTIBKSGBDQIKGRIRKSNAGSKKKIKIPVVAAPAD
 KCTPGCSVAQPDVYCSNDICILHAAATMRPLSGKOKTKPKKTKPKSLP
 KCSVVGITKISSVHKLASKEKENPYKATMLASRSTSGKRAKESSTSMASDHYN
 AVKPKPKPTALSPILLSKCTHPKAGPGPSHHLGCLGLSRTRVLVLIVASS
 SIAPRSRYDASGPQVFLPSLMSLSGMLFKSCVGLMLRAISYFSRPM"

ORIGIN

Query Match 39.3%; Score 1025.2; DB 11; Length 4619;
 Best Local Similarity 71.9%; Pred. No. 1,4e-243;
 Matches 1495; Conservative 0; Mismatches 533; Indels 50; Gaps 10;
 21 TCCGCGGCTTCGGGGAATGCTGCGAGACCTTAGAGGCTTCGAGCTTACCAACG 80
 68 TCCGTGTAGCTCTGGAATGCTGCGAGATCCCGCGGCGGAGCTTTTAAGAG 127
 81 GAACAGCTCTAGATATCTGAGTTTGAATAACAGACCTGTACTCGTAACAGTG 140
 128 CAGTCCCACTGCTTGGTGTTCGAAGCTCCGAAATCTTCTCATTTGATGAACGTG 187
 141 GCTGACACAGTGTGTGTGAGCTCTGCTTGTGACCCAGAGGTTTGTCTGCC 200
 188 ACTGACAAAGTGGGTGAG-GCTTGCCCTGCTGTAC-----TGCCC 233
 201 AGGGTTTGTGTATTAGATT-CAGGGAAGGTGCAAGCTTTCAGGTTGAG 259
 234 CAGGTTTAAATTTTATGAGATTTCAGCCAAAGTTTCAAGCTTTCAGTTTGGGA 293
 260 CAGGTATGACCAAAAGCGAAGCCGAGCAATGAGAGCACTTAAGCCATCAAAACCA 319
 294 CAGGTATGATATTAAGGCACTGAGCAATGAGAGCAACCAAGGCTATCAAAACCA 353
 320 CAGCAAAAGATTGAGAAACATGGGCTTTTCGAGAGCACTATTCGCAACGAGAG 379

Db 354 CAGGTATGAGATTTCAGAAACCTGGGCTTTTCGAGAGCACTATTCGCAACGAGAG 413
 380 GCGCAGAGGAGCGCGAGGCTGACCTCACTGAGCCGCAACCCCAACAGCAGAGCTGGGCC 439
 414 GTGCAAGAGACACGAGAGGTGAGACCCCAAGAGAGCAAC-----ACAGGACATTAAC 467
 440 TGTCCCTGCGGCGAGTGGAGAGCGCCCAAGGCACTGAGGCGTGGAGCAGTTCTGA 499
 468 TCTCCCTGCGGCGAGTGGAGAGCGCCCAAGGCACTGAGGCGTGGAGCAGTTCTGA 527
 500 CATTGCGGCGCGCGCGCAGAGAGAGAGCTGCTTCTCTGAGAGATTCTGTTGAGC 559
 528 CCAAGTGTGGCGCCGAGAGGAAAGATGTGCGGTGTCTCTGAGAGATTCTGTTGAGC 587
 560 CCAAGTGTGGCGCCGAGAGGAAAGATGTGCGGTGTCTCTGAGAGATTCTGTTGAGC 619
 588 CCAAGTGTGGCGCCGAGAGGAAAGATGTGCGGTGTCTCTGAGAGATTCTGTTGAGC 647
 620 CTGAGACCAAGAGCGGCGCCGAGAGGAAAGATGTGCGGTGTCTCTGAGAGATTCTGTTGAGC 679
 648 CTGAGATCAGAAAGTGGCGCTGTATCTGACTCTTA--GGAAAGACATCTGCTCTT 704
 680 CTGAAAAGGTGAAAGAGGAGATGACCAAGATGACCTCCGATAGTGAACAGGATGCGC 739
 705 CTGAAAAGGTGAAAGAGGAGATGACCAAGATGACCTCCGATAGTGAACAGGATGCGC 764
 740 TGAAGTGAAGAGCTTCAAGATGCTTTCGAGAGAGCGGAAACAGAGCCCACTGAGA 799
 765 TTAAGTTGAAGAACTTCAAGACCGCTTCGAGAGAGCGAGAGCAAGAACTGTGAGA 824
 800 GGGCCCTGAAAGAGATCCAGATGCGCTCGAGAGAGAGCGCGGAGAGAGGCTCCGCG 859
 825 GGTCCCTGAGAGCAATCAGAAATGCTGAGAGAGAGAGCGCAGAGAGAAATTTCTGCG 884
 860 AGACTGTGAGCTCCGAGCGCAGTGAACATGTTGAGAGGCGCTCTGCCAGTAAACAGAGC 919
 885 AAATGAGATGCTCAAAATAGAGAGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGC 944
 920 CCGAGAACATCAAGGAGGTTGTGTCCAGGCTGAGAGAGTGAACAGAGAGTGAATGG 979
 945 CTGAGGCTAGTCAAGGAGCAAGTGTCTCCAGTCAAGAGCAGATGACATTAAGAAATCAATTGG 1004
 980 AGGGAAGGCGCTCGAGACATCAAGATGAGAGGCTGAGAGCTTGGGCGACCGAAGC 1039
 1005 AAGGGAAGGCGACTCGAGAAATACAGAGAGAAACCCAGGGAAGCGGAGAAACCAAGC 1064
 1040 CTGAATGTGAGGTTACAGACCCCAAGCGCTGTATTGCAATTTGCGCAGAGCTTCAACA 1099
 1065 CTGAGTGTAGGTTTACAGACCCCAAGCGCTGTATTGCAATTTGCGCAGAGCTTCAACA 1124
 1125 ACAGGTTTATGATCTCTGTGATGGGTGAGAGAGTGTCTCAATGTGATCTGTGTGGA 1184
 1160 TTTCTGAGGCTGAGAGAGCTTTTGAAGAGAAATGGGAAAGCTATATCTGCCAAAT 1219
 1185 TTTCTGAGGCTGAGAGAGCTTTTGAAGAGAAATGGGAAAGCTATATCTGCCAAAT 1244
 1220 GCAACATTTTTCGAGAGGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1279
 1245 GCAACATTTTTCGAGAGGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1304
 1280 AATGAGACCTGAGAGTCTGATGGAACGATTTTCAAGTATAGGAACAATAGAGAGA 1339
 1305 GGTGAGATCTGTGGGTGATGGAACGATTTTCAAGTATAGGAACAATAGAGAGA 1364
 1340 AGTCTAG 1399
 1365 AGTCTAG 1424
 1400 AGAAGAACTCAAGATCTTTCAGAGCTGTGATAGAGAGCGCTGTGCTCAAAATGATTTG 1459

Db 1425 AGAAAACTCAGATATTCAGACCTGTGTAGAGGCTCTGTGCTCTCTAATTCATTG 1484
 QY 1460 GCCCGGTCGTCTCACTGGGCGCAGCCCGACTCGGTACTGACGATATGCTATTC 1519
 Db 1485 GCCCTGGTTCAGAGTACGACAGCCCTGACTCTGTATATGCGATATGACTGATTC 1544
 QY 1520 TCAACACGCGCGCAGCGCAATGAAGTTTCTAAGCTCAGTTAAAGACAGAACCAAGC 1579
 Db 1545 TCAACACGCGCAGCTACATGAGATTTCTAAGTTCAAGTTAAAGACCAAAACCAAAAC 1604
 QY 1580 CTAAAGAAAAGATGAAGTGAAGCCAGAACCCAGTCTTCCGAAATGCGGTCTCAG 1639
 Db 1605 CCAAGGAAAAGCTCAAGACGAGCAAGAAAGTTCAAGTCTTCAAAATGCAAGTTCAG 1664
 QY 1640 CAGGATTTAAATCTTCTGTGCAAGAGACCAAGTCCAGAAAAAAAGAGACACAG 1699
 Db 1665 TGGGATTTAAATCTTCTGTGCAAGAGACCAAGTCCAGAAAAAAAGAGAAAAACCCAG 1724
 QY 1700 TGAAGAAAGCAGTGTGTCTCTGCGCAGAGTGAAGCACTCGGAAAGAGAGAGCTTGTG 1759
 Db 1725 TGAAGAA---AGTATGTCTGGCTTCAGAGTGAAGACTTCTGGAAAGAGAGAGCTGTG 1781
 QY 1760 AGAGCAGACCCCGTCTGTGGCGAGCGATCACAATTAATGACGATTAAGCCAGAAAA 1819
 Db 1782 AGAGCAGACCACTCTCTGGGCAAGTGAACAACACTCAATGCTGTGAAGCGAGAAAGC 1841
 QY 1820 CTG-----CTGCTCCCTGCGCGCTCACTGTGTATTAATGATATGATCACTTACGAG 1870
 Db 1842 CAGAGAAAGCCCACTGCACTCTGCGCCACTTATTAAGTAAATGATGATCACTCCAAAG 1901
 QY 1871 TTGGCTCTCTGAGCCCTCTCCGTTCTTTCTGATAGCAATCCCTGGGCTGTTCAGAG 1930
 Db 1902 CTGCTCTCCAGGCCCCCTCCATCATCT---GGGTGGCTGCTGTGGGCTGTCTAGAG 1956
 QY 1931 TGGAGTTGCAAGCTTTGTGTTAAGTATCAAGACACCGGTGCAACATCAGCCGGAAG 1990
 Db 1957 CAG-----AGTCTGGGTGTCTGTGTGATGATGAGCAAGAG---TCACTCCAGC 2007
 QY 1991 CAGAGCCCATGTCAGATGAGTGCCTCTGTGCTCTGTGTCATCCTTATGTTTCAGAGT 2050
 Db 2008 CAGAGCAGATACCAAGATGCTCTGAGACCCAGAGTGTCTGCTTACCTGTGAGGCT 2067
 QY 2051 TCCTGTCACTGTTTCCAAAGCTGTAACTCACTGCT 2088
 Db 2068 CTCTGGGTGTCTCTAAAGAGCTGTGTAGGCTCATGT 2105
 RESULT 3
 AL578222/c 1086 bp mRNA linear EST 01-JUN-2003
 LOCUS AL578222 Homo sapiens HELA CELLS COR 25-NORMALIZED Homo sapiens
 DEFINITION cDNA clone CS0DK002YJ13 3-PRIME, mRNA sequence.
 ACCESSION AL578222
 VERSION AL578222.2 GI:31316440
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1086)
 L1.M.B., Gruber,C., Jesse,J. and Polayes,D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On Feb 16, 2001 this sequence version replaced gi:12942093.
 Contact: Genoscope - Centre National de Sequencage
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 9799.r for
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0DK002CE07NP1cluster=9799.r. Contact :

Peng Liang Email : liang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Fareaday Avenue Genoscope sequence ID : CS0DK002CE07NP1.
 Location/Qualifiers
 FEATURES
 source
 1..1086
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DK002YJ13"
 /cell_type="HELA CELLS COR 25-NORMALIZED"
 /clone_lib="HELA"
 /clone_lib="Homo sapiens HELA CELLS COR 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."
 ORIGIN
 Query Match 38.6%; Score 1007.2; DB 9; Length 1086;
 Best Local Similarity 97.5%; Pred. No. 2.8e-239;
 Matches 1047; Conservative 5; Mismatches 18; Indels 4; Gaps 3;
 QY 1047 TGAAGTTACACACCCCAACGCCCTGATTCATTGGCCCGCAGCTCAACAACAGGTT 1106
 Db 1072 TGTGAGTTACGACCCCAACGCCCTGTAT--GCTTGCSCGAGCTCAACAACA--GTT 1016
 QY 1107 TATGATTTCTGTGACCGCTGTGAAGAAATGTTTCAATGCGATGTGTGCGCATTTCTGA 1166
 Db 1015 TATGATTTCTGTGACCGCTGTGAAGAAATGTTTCAATGCGATGTGTGCGCATTTCTGA 956
 QY 1167 GGCCTGAGGAGCGCTTTTGAAGAAATGCGAAGACTATATCTGCCAAATGCACTCAAT 1226
 Db 955 GGCCTGAGGAGCGCTTTTGAAGAAATGCGAAGACTATATCTGCCAAATGCACTCAAT 896
 QY 1227 TCTGCAAGTGCAGATGAGATCTATTCAAGAACGCGATGCGAGAGCTTAATGAGAG 1286
 Db 895 TCTGCAAGTGCAGATGAGATCTATTCAAGAACGCGATGCGAGAGCTTAATGAGAG 836
 QY 1287 ACCTGAGATGCTGTGAGCAGATGCGATGTTGTAAGATAGAGAAATGAGAGAGTCTAG 1346
 Db 835 ACTGAGATGCTGTGAGCAGATGCGATGTTGTAAGATAGAGAAATGAGAGAGTCTAG 776
 QY 1347 CGAAGACCAAGGATTAAGGTTAGATTAAGAAAGCTGCAATTCAGTGGCAAGAA 1406
 Db 775 CGAAGACCAAGGATTAAGGTTAGATTAAGAAAGCTGCAATTCAGTGGCAAGAA 716
 QY 1407 ACTCAAGATCTTCCAGCTGTATATAGAGGCGCTGTGCTCAAAATGATTTGGCCCCGG 1466
 Db 715 ACTCAAGATCTTCCAGCTGTATATAGAGGCGCTGTGCTCAAAATGATTTGGCCCCGG 656
 QY 1467 GTGCTGTACAGTGGCGAGCCCGACTCGGTGTACTGCAATGATGATCTCTCAACA 1526
 Db 655 GTGCTGTACAGTGGCGAGCCCGACTCGGTGTACTGCAATGATGATCTCTCAACA 596
 QY 1527 CGCGCAGCGCAATGAAGTTTCTAAGCTCAGGTAAAGACAGAACCCAAAGCTTAAGA 1586
 Db 595 CGCGCAGCGCAATGAAGTTTCTAAGCTCAGGTAAAGACAGAACCCAAAGCTTAAGA 536
 QY 1587 AAAGATGAAGATGAAGCCAGAGAACCCAGTCTTCCGAAATCGGTGCTCAAGCAGTAT 1646
 Db 535 AAAGATGAAGATGAAGCCAGAGAACCCAGTCTTCCGAAATCGGTGCTCAAGCAGTAT 476
 QY 1647 TAAATCTCTTGTGCAACAAGACAGCTCAGAAAAAAAGAGACCAAGTGAAGAA 1706
 Db 475 TAAATCTCTTGTGCAACAAGACAGCTCAGAAAAAAAGAGACCAAGTGAAGAA 416
 QY 1707 GCGAGTGTGTCTCTGCGCGAGTGAAGCACTCGGAAAGAGACGCTTGTGAGAGCAG 1766
 Db 415 GCGAGTGTGTCTCTGCGCGAGTGAAGCACTCGGAAAGAGAGCAGCTTGTGAGAGCAG 356
 QY 1767 CACGCGCTGTGGGAGAGGATCAATTAAGATGCAATTAAGCCAGAAAGACGCTGCG 1826
 Db 355 CACGCGCTGTGGGAGAGGATCAATTAAGATGCAATTAAGCCAGAAAGACGCTGCG 296

QY 1827 TCCTCCGCGCTGACTGTTGTATTAATGATATGATACCTTAGGGCTTGGCTCTCCGAGACC 1886
 DB 295 TCCCTCCGCGCTGACTGTTGTATTAATGATATGATACCTTAGGGCTTGGCTCTCCGAGACC 236
 QY 1887 CTCCGCTGCTTCTGATAGGACCATCCCTGGGCTGTCACAGACTGAGAGTTGACAGCTTT 1946
 DB 235 CTCCGCTGCTTCTGATAGGACCATCCCTGGGCTGTCACAGACTGAGAGTTGACAGCTTT 176
 QY 1947 GTGTTAAGCTGATACAGACACCGGCTGACACCATCAGCGGAGAGAGAGCCCATGTCAG 2006
 DB 175 GTGTTAAGCTGATACAGACACCGGCTGACACCATCAGCGGAGAGAGAGCCCATGTCAG 116
 QY 2007 GATGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2066
 DB 115 GATGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 56
 QY 2067 CAAGCTGTAACCTC-ACGCTGTAACCTC-ACCTTAATGATGATGCTTAAAT 2119
 DB 55 CAAGCTGTAACCTC-ACGCTGTAACCTC-ACCTTAATGATGATGCTTAAAT 2
 RESULT 4
 BC060249
 LOCUS
 DEFINITION Mus musculus death inducer-obiterator 1, mRNA (cDNA clone
 IMAGE:5826043), containing frame-shift errors.
 ACCESSION BC060249
 VERSION BC060249.1 GI:37805411
 KEYWORDS
 SOURCE HTC.
 ORGANISM Mus musculus (house mouse)
 REFERENCE
 AUTHORS Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 4651)
 Strauberg, R.D., Collins, P.S., Wagner, L.H., Derge, J.G.,
 Kluemper, R.D., Collins, P.S., Wagner, L.H., Derge, J.G.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Wax, S.I., Wang, J., Hsieh, F.,
 Datchenko, L., Maruina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stopleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Schaefer, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S.S., Loguercio, N.A., Peters, G.J.,
 Abramson, R.D., Molihany, S.J., Bosak, S.A., McEwan, P.J.,
 McEwan, R.D., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Wolter, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hilyk, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Hellon, E., Kettleman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmitt, J., Myers, R.M.,
 Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Smalls, D.E.,
 Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 22388257
 12477932
 2 (bases 1 to 4651)
 Strauberg, R.
 Direct Submission
 Submitted (20-OCT-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgabde-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library Preparation: M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada

info@bcsc.bc.ca
 Steven Jones, Jennifer Asano, Ian Bosdel, Yaron Butterfield,
 Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Leticia Hsiao, Martin Krzywicki, Reta Kutsche, Oliver Lee, Soo
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
 Nee, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
 Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
 Michael Thorne, Miranada Teai, Nastasia van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.
 Clome distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAX Plate: 126 Row: K Column: 20
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, similarity but not identity to protein
 This clone has the following problem: frame shifted.
 location/Qualifiers
 1..4651
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:5826043"
 /tissue_type="Brain, mouse, 13.5,14.5,16.5,17.5 dpc"
 /clone_id="NIH BMAP_FY0"
 /lab_host="DH10B"
 /note="Vector: pIX-ASC"
 ORIGIN
 Query Match 38.4%; Score 1002.6; DB 11; Length 4651;
 Best Local Similarity 73.5%; Pred. No. 5.8e-238;
 Matches 1414; Conservative 0; Mismatches 474; Indels 37; Gaps 9;
 QY 175 GCTTGAACCCAGAGGTTGCTGTCGACGAGTTTGTGTTGTTAGATT-CAAGGAA 233
 DB 187 GCTTGAACCCAGAGGTTGCTGTCGACGAGTTTGTGTTGTTAGATT-CAAGGAA 246
 QY 224 AAGTGTCCAGCTTTCAGTGTGAGCAGTATGACACAAAGGACCCGAGCAATGA 293
 DB 247 AAGTGTCCAGCTTTCAGTGTGAGCAGTATGAGTATGAGGACCTGAGCATGA 306
 QY 294 GGAGGACCTTAAGGCTTCAACCAACCAAGGTTTGTGTTGTTAGATT-CAAGGAA 353
 DB 307 GGAGGACCTTAAGGCTTCAACCAACCAAGGTTTGTGTTGTTAGATT-CAAGGAA 366
 QY 354 AAGGACCACTATGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 413
 DB 367 AAGGACCACTATGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426
 QY 414 GCCATCCCCCAG 473
 DB 427 GCAAC-----ACGACAGCATATACCTTCTCCCTGCGCGCATGAGACGCAACAAAG 480
 QY 474 CACTGAGCGGAG 533
 DB 481 TACTGAGAGGTTAAG 540
 QY 534 TGTCTCCCTGAG 593
 DB 541 GGTGTCCCTGAG 600
 QY 594 CTCCGAGGCGAGGTTGAG 653
 DB 601 TTCCGAGGCGAGGTTGAG 660
 QY 654 AGCTGTGAAG 713
 DB 661 A---GGAAAG 717
 QY 714 CACCTCCATAGTGAACAG 773
 DB 718 CACCTCCATAGTGAACAG 777

Oy	774	GAACGGGGAAACAAGAACCCCACTGAGAGGCCCTCTAAAGGGATTCAGAGTGGCCCTGGCGGAA	833
Db	778	AAACCGAGAGCAAGAACCTGTGGAGAGGATCTCTGAGAGGCACTGAGAAATCCCTGAGGAA	837
Oy	834	GAACGGCCGGAGAGAGGGTCCCGCGAGACTGTGGGCTCCGAGGCCAGTGTGACACTGTGGA	893
Db	838	GAACCGCAGAGAGGAAGATTCTGGCGAAACTGTGGAGTGTCCAAATATGGCAAGTGCAGCA	897
Oy	894	GGCGCTCTGCCCACTAAGCAGAGAGCCCGAAGACGATCAGGGGGTGTGTCTCCAGGCTGG	953
Db	898	GGACAGACCTTCTGTATAGCAGAGGCTCGAGGCTAGTCAAGGACCAAGTGTCCCACTGACA	957
Oy	954	GAAGAATGACAGAGAGATAGTTGAGAGGAAAGCGGCTCAGGACATCAAGAATGAGGA	1013
Db	958	GACAGATACATAGAAATCAGTTGTGAAGGAGAGGCCACTAGAGGAAATACAGAGGAAAA	1017
Oy	1014	GCCCTGAGACTTGGGCGCAGCCGAAGCCCTGAATTGTAGGGTTACGACCCCAAGCCCTGTGTA	1073
Db	1018	CCCCAGGGAGACGGGGCAAAACCAAGGCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTGTA	1077
Oy	1074	TTGCAATTTGCCCCCAGCCTCACAACAAGTTTATGATTTGTCTGTGACCGCTGTGAGA	1133
Db	1078	CTGCACTTGGCCGCGACGCTCAACAACAGGTTTATGATCTGTGTATCGGTGTGAGGA	1137
Oy	1134	ATGGTTTCATGGCGGATTTGTGTGGGCAATTTGAGGCTGAGGAGAGGCTTTTGGAAAGGA	1193
Db	1138	GTGGTTCCATGTGTATCTGTGTGGATTTCTGAGGCGCGAGGGCGGCTCTGTGAAGGAA	1197
Oy	1194	TGGGAGAGCTATATCTTGGCCCAACTGCAACATTTGTGCAATGTGAGATGAGACTATTC	1253
Db	1198	CGGGAGAACATCAATCTGGCCCAAAATTTGACCAATTTTGCAAGTGCAGATGAGACAAAGG	1257
Oy	1254	AGAAACGGCAGATCAGCAGAGAGCTAAATGAGACCTTGAGATGTCTGATGGCACCGATTG	1313
Db	1258	TAGCCCAACCGATGAGCAGAGACTCTGGGTGCGAGATCTGTGGTGTGATTTGGACAGACTG	1317
Oy	1314	TACAAATATAGAACATATGAGCAGAGATCTAGCGAAGACCAAGGGATTAAGGGTGAAT	1373
Db	1318	CACAAGCAATAGGAGCAGTGTAGCAGAAATCCGAGAAACCAAGGCGATTAAGGGTGAAT	1377
Oy	1374	TGAGAAACCTCAAAATCCAGTGGCAAGAAAGTCAAGATCTTCCAGCTGTGATGGA	1433
Db	1378	TGAGAAAGCACAAACCCACGAGGCAAGAAAATTCAGAGATTTCCAGCTGTGTGTGA	1437
Oy	1434	GAGCGCTGTGCTCTCAAAATGTATTTGGCCCGGGGTCTGTCAAGTGGCGAGCCCGACTC	1493
Db	1438	GGCTCTGTGTCTCTTAATATGCAATTTGGCCCTGGGGTTCCAAGTGTACACAGCTGTGACTC	1497
Oy	1494	GGTGTATCTGCAGTATATGACTGTATCTCAAAACAGCCCGCACAGCAATGAAGTTCTTAAG	1553
Db	1498	TGTGTATTGCAGTATATGACTGTATCTCAAAACAGCGACAGCTATACATGAGATTTCTTAAG	1557
Oy	1554	CTCAGGTAAAGAAC-AGAAAGCCAAAGCTTAAAGAAAATGAAAGATGAAAGCCAGAGAAC	1612
Db	1558	TTCAAGTAAAGAACAAAACCAAAACCAAGAGAAAGTCAAGCGAAGCCAGAAAGT	1617
Oy	1613	CCAGTCTTCCGAATGCGGTGCTCAGCGAGGTATTAATATCTTCTGTGTCAAAAGAAC	1672
Db	1618	TCAGTCTTCCAAAATGCAAGTGTTCAGGTGGGGATTAAATCTCTTCTGTGTCAAAAGAAC	1677
Oy	1673	CAGCTCCAGAAAAAAGAGACCAACAGTGAAGAAAGCAGTGGTGTCCCTGCGCGAGATG	1732
Db	1678	TAGCTTCAGAGAAAAGGAAAAACCAAGTGAAGAA--AGTATGTGGCTTCCAGAGATG	1734
Oy	1733	AAGCACTGGGAAGGAACAGCTGTGAGACAGCAGCGCTGTGGGCGAGCATCA	1792
Db	1735	AGACTTCTGGGAAGGAGCAGCTGTGAGACAGACACATCTCGGCAAGTGAACACA	1794
Oy	1793	ATTCAATGCAATTAAGCCAGAAAAGACTG-----CTGGTCCCTGGCGTGTCACTGT	1843
Db	1795	ACTCAATATCTGTAAAGCCAGAGAAAGCCAGAGAACCCCACTGTCACTCTGGCCCACTTAT	1854

QY	1844	TGATTAATATTAATATACACTTGGGGTTGGCCCTCCGAGACCCCTCCGGTCTTCTGGA	1903
Db	1855	TGAATTAATATGACTATACACCAAGAGCTGGCTTCCAGGCCCTCCCATCATCT-----	1909
QY	1904	TAGCATCCCCCTGGGCCCTGTCACAGACTGGAGATTGACGCTTTGTGTTAAGCTATACA	1963
Db	1910	GGGATGGCTGCCTGGGGCTGTCTTAGGACAG-----AGTCCCTGGGTGTTCTGGTGTCTATA	1964
QY	1964	GACACCGGCTGACACATCAAGCGGAGACAGGCCATGTCCAGGATGCTCTCTCTGCGCC	2023
Db	1965	GTACCGACGACGAC-----TACCTGCGACCGACAGACGATACCAAGATGCTCTCTGACCCCA	2020
QY	2024	TGATTCATCCCTTAAGTCTGTCAAGACTTCTCTGTCACTGTTTTCCAAAGCTGTAACCTCA	2083
Db	2021	GGATTTCTCTGCTAGCTCTGTGAGCTCTCTGTGGGTGTTCTTAAGAGCTGTGTAGGCTT	2080
QY	2084	CTGCT 2088	
Db	2081	CATGT 2085	

RESULT 5

LOCUS	AL555810	1093 bp	mRNA	linear	EST 31-MAY-2003
DEFINITION	AL555810 Homo sapiens HELA CELLS	COT 25-NORMALIZED	Homo sapiens		
	cDNA clone CS0DK002YJ13 5-PRIME, mRNA sequence.				
ACCESSION	AL555810				
VERSION	AL555810.2	GI:1277616			
KEYWORDS	EST.				
SOURCE	Homo sapiens	(human)			
ORGANISM	Homo sapiens				

REFERENCE

REFERENCE 1 (bases 1 to 1093)
AUTHORS I. W. B., Gruber, C., Jesses, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 15, 2001 this sequence version replaced gi:12897895

BP 191 91006 EVRY cedex - France
Email: seq@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9799.1 For
more information about this cluster, see

Reng Liang Email : liang@liverectn.com UKL :
<http://fulllength.invitrogen.com/InvitrogenCorporation1600>
 Parady Avenue Genoscope sequence ID : CSDBK002CB07QP1.

Source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK002YJ13"
/cell_type="HELA CELLS COT 25-NORMALIZED"

```

ORIGIN

Query Match	36.7%	Score 957.2	DB 9	Length 1093
Best Local Similarity	97.8%	Pred. No. 7.8e-22		
Matches 1011; Conservative	3	Mismatches 11	Indels 9	Gaps 4

41 GGCCTCCAGACCTTAGAGGCTGGG-----GAGCTTACTCCAGGGAAACAGCCTCTAGAT 95
 42 |||||
 43 |||||
 44 |||||
 45 |||||
 46 |||||
 47 |||||
 48 |||||
 49 |||||
 50 |||||
 51 |||||
 52 |||||
 53 |||||
 54 |||||
 55 |||||
 56 |||||
 57 GGAATCCAGACCTTAGAGGCTGGGCTCCAGGAATTACTCCAGGGAAACAGCCTCTAGAT 116
 58 |||||

QY 96 AATCTGAGTTGTTAAATACGAAGCCCTGTTACTGTGACAGTGGCTGACACAGCTTT 155
 DB 117 AATCTGAGTTGTTAAATACGAAGCCCTGTTACTGTGACAGTGGCTGACACAGCTTT 176
 QY 156 GTTGTGAGCCCTGGCTGTGCTGTGACCCAGAGGTTTGTCTGCGAGGGTTTGTGTTGT 215
 DB 177 GTTGTGAGCCCTGGCTGTGCTGTGACCCAGAGGTTTGTCTGCGAGGGTTTGTGTTGT 236
 QY 216 ATTTAGATTTTCAAGGAAAAGTGTCCAGCTTTCAAGTGTGAGAGAGGTTATGACGACAA 275
 DB 237 ATTTAGATTTTCAAGGAAAAGTGTCCAGCTTTCAAGTGTGAGAGAGGTTATGACGACAA 296
 QY 276 AGCGGACCCGAGCAATGAGAGGCACTTAAGGCTTCAAACTCCACAGCAAAAGATTGAG 335
 DB 297 AGCGGACCCGAGCAATGAGAGGCACTTAAGGCTTCAAACTCCACAGCAAAAGATTGAG 356
 QY 336 GAAAAATGAGGGTTTTGAGAGGACCACTATGCGCCAAAGAGAGAGGCGGAGGAGCGCGA 395
 DB 357 GAAAAATGAGGGTTTTGAGAGGACCACTATGCGCCAAAGAGAGAGGCGGAGGAGCGCGA 416
 QY 396 GAGCTGACCCACTGAGAGCCGACCCCAAGAGAGAGCTGGGCTGTCTCTGCGCGCAG 455
 DB 417 GAGCTGACCCACTGAGAGCCGACCCCAAGAGAGAGCTGGGCTGTCTCTGCGCGCAG 476
 QY 456 TGGAGAGGAGCCCAAGGCGCACTGAGAGGCTGTGAGAGAGTTCTGACATTTGGCGGCGCG 515
 DB 477 TGGAGAGGAGCCCAAGGCGCACTGAGAGGCTGTGAGAGAGTTCTGACATTTGGCGGCGCG 536
 QY 516 CGGAGAGAGAGAGAGCTGTCTCTCTGAGAGATTTCTGAGAGCCAGCTCTGCGCGCG 575
 DB 537 CGGAGAGAGAGAGAGCTGTCTCTCTGAGAGATTTCTGAGAGCCAGCTCTGCGCGCG 596
 QY 576 CACAGACCCGAGAGAGAGCTCTGAGAGGAGAGCTGTGAGAGAGCTTTGAGAGCAGAGCGG 635
 DB 597 CACAGACCCGAGAGAGAGCTCTGAGAGGAGAGCTGTGAGAGAGCTTTGAGAGCAGAGCGG 656
 QY 636 CCCCCAGTCTGTTTCAAGCTGTGAGAGAGAGAGAGAGCTTTTGTGAGAGAGTGAAGG 695
 DB 657 CCCCCAGTCTGTTTCAAGCTGTGAGAGAGAGAGAGAGCTTTTGTGAGAGAGTGAAGG 716
 QY 696 AGGGAGTACCAAGATGACCACTCTGATAGTACAGAGAGAGCTGACCTTGAAGAGCT 755
 DB 717 AGGGAGTACCAAGATGACCACTCTGATAGTACAGAGAGAGCTGACCTTGAAGAGCT 776
 QY 756 TCAGATTCGCTTTCAGAGAGAGCGGAAACAGAGAGCCACTGAGAGAGCCCTTGAAGAGAT 815
 DB 777 TCAGATTCGCTTTCAGAGAGAGCGGAAACAGAGAGCCACTGAGAGAGCCCTTGAAGAGAT 834
 QY 816 CCAGAGTCCCTGCGAGAGAGAGCGGAGAGAGAGTCCCGCCAGAGAGCTGTGGCTCCGA 875
 DB 835 CCAGAGTCCCTGCGAGAGAGAGCGGAGAGAGAGTCCCGCCAGAGAGCTGTGGCTCCGA 894
 QY 876 GGGCCAGTACACTGTGAGAGAGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 934
 DB 895 GGGCCAGTACACTGTGAGAGAGAGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 954
 QY 935 GGGTGTGTCTCCAGAGCTGGAG 994
 DB 955 GGGTGTGTCTCCAGAGCTGGAG 1014
 QY 995 AGGAGATCAAAAGATGAG 1054
 DB 1015 AGGAGATCAAAAGATGAG 1073
 QY 1055 AGGAG 1087
 DB 1074 AGGAG

RESULT 6
 AL580669 1033 bp mRNA linear EST 01-JUN-2003
 LOCUS AL580669 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
 DEFINITION

Homo sapiens cDNA clone CS0DJ007YD06 5-PRIME, mRNA sequence.
 AL580669
 VERSION AL580669.2 GI:31318936
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 REFERENCES
 AUTHORS Li, W.B., Gruber, C., Jeessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT On Feb 15, 2001 this sequence version replaced gi:12946913.
 Contact: Genoscope - Centre National de Sequencage
 Genoscope - BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 9799.r For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0DJ007YD0630P1&cluster=9799.r. Contact :
 Feng Huang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0DJ007YD0630P1.

FEATURES

source
 Location/Qualifiers
 1..1033
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DJ007YD06"
 /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
 /clone_11b="Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 34.9%; Score 910.6; DB 9; Length 1033;
 Best Local Similarity 97.0%; Pred. No. 3,1e-215;
 Matches 934; Conservative 14; Mismatches 9; Indels 6; Gaps 2;
 QY 33 GGGGAAATGCTGCGAGAGACCTTGAAGGCTGCGAG-----CTTACTCAAGGAAACAGC 87
 DB 67 GGGGAAATGCTGCGAGAGACCTTGAAGGCTGCGAGCTTACTCAAGGAAACAGC 126
 QY 88 CTCTAGATTAATCTGAGTGTGTAATAATCGAAGCCTGTACTCGTGAACAGTGGCTGACA 147
 DB 127 CTCTAGATTAATCTGAGTGTGTAATAATCGAAGCCTGTACTCGTGAACAGTGGCTGACA 186
 QY 148 ACAAGTGTGTTGAGAGCTGAGCTGTCTGCTTGAACCCAGAGGTTTCTGTGCCAGGGTTT 207
 DB 187 ACAAGTGTGTTGAGAGCTGAGCTGTCTGCTTGAACCCAGAGGTTTCTGTGCCAGGGTTT 246
 QY 208 TTGTTGTATTTAGATTTCAAGGAAAGTGTCCAAAGCTTTCAAGTGTGAGAGAGATAG 267
 DB 247 TTGTTGTATTTAGATTTCAAGGAAAGTGTCCAAAGCTTTCAAGTGTGAGAGAGATAG 306
 QY 268 GACGACAAAGGAGACCCGAGCAATGAGAGAGGACCTTAAGGACCTCAAAACCCACAGCAAA 327
 DB 307 GACGACAAAGGAGACCCGAGCAATGAGAGAGGACCTTAAGGACCTCAAAACCCACAGCAAA 366
 QY 328 GAGTTCAAGGAAACATGAGGGTTTTCAGAGAGCACTATGCCAAGGAGAGAGGCGGAGGG 387
 DB 367 GAGTTCAAGGAAACATGAGGGTTTTCAGAGAGCACTATGCCAAGGAGAGAGGCGGAGGG 426
 QY 388 GACGCGAGGCTGACCCACTGAGAGCCGCCACCCCAACAGACAGAGTGGGCTGTCTCTG 447
 DB 427 GACGCGAGGCTGACCCACTGAGAGCCGCCACCCCAACAGACAGAGTGGGCTGTCTCTG 486

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9799.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CL0BB005ZB1RP1&cluster=9799.r. Contact :

Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CL0BB005ZB1RP1.

FEATURES

source

1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CL0BB005ZB1"
/issue_type="NEUROBLASTOMA"
/clone_1ib="Homo sapiens NEUROBLASTOMA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN

Query Match 33.1%; Score 868; DB 9; Length 1201;
Best Local Similarity 95.1%; Pred. No. 1.3e-204;
Matches 919; Conservative 7; Mismatches 37; Indels 3; Gaps 3;

65 GAGACTTATCCACGCGGAAACGCTTAAATATCTGAGTTGTTGAAATACGAGCTTG 124
160 GAGACTTATCCACGCGGAAACGCTTAAATATCTGAGTTGTTGAAATACGAGCTTG 219
125 TTAAGTGTGACAGGCTGTGACAAAGTGTGTGTGAGCTGTGCTGTGAGCC 184
220 TTAAGTGTGACAGGCTGTGACAAAGTGTGTGTGAGCTGTGCTGTGAGCC 279
185 AGAGGTTTGTGTCGACGAGGTTTGTGTGTATTTAGATTTCAGGAAAGTGTCAAG 244
280 AGAGGTTTGTGTCGACGAGGTTTGTGTGTATTTAGATTTCAGGAAAGTGTCAAG 339
245 CTTCAGTTGTGAGAGCGATATGAGACAAAGCGACCTGAGCAATGAGAGGCACTTA 304
340 CTTCAGTTGTGAGAGCGATATGAGACAAAGCGACCTGAGCAATGAGAGGCACTTA 399
305 AGGCAATCAAAACCCACAGCAAGATTCAGGAAATCATGGGTTTTCGAAAGCACTTA 364
400 AGGCAATCAAAACCCACAGCAAGATTCAGGAAATCATGGGTTTTCGAAAGCACTTA 459
365 TCGCCAAAGGAGGCGGCGAGGCGGAGGCTGACCTGAGAGCGCCACCCCAAC 424
460 TCGCCAAAGGAGGCGGCGAGGCGGAGGCTGACCTGAGAGCGCCACCCCAAC 519
425 AGCAGAGCTGGGCGCTGTCTCCGCGGCACTGAGAGCGCCCAAGCGCG 484
520 AGCAGAGCTGGGCGCTGTCTCCGCGGCACTGAGAGCGCCCAAGCGCG 579
485 TGAAGAGATTCTTGAACATTGGCGGCGCGCGGAGGAGGAGCATGCTCTCCCTG 544
580 TGAAGAGATTCTTGAACATTGGCGGCGCGCGGAGGAGGAGCATGCTCTCCCTG 639
545 AGGATTTCTGTGAGAGCCAGCTCTGCGCGGCAAGAGCCGAGAGAGCTCCGAGGCA 604
640 AGGATTTCTGTGAGAGCCAGCTCTGCGCGGCAAGAGCCGAGAGAGCTCCGAGGCA 699
605 GGTGTGAAAGCGCTTCTGAGACCAAGAGCGGCGGCGGAGTGTCTTCAAGCTGTGAAG 664
700 GGTGTGAAAGCGCTTCTGAGACCAAGAGCGGCGGCGGAGTGTCTTCAAGCTGTGAAG 759
665 AACGACCAAGCTCTTCTGAAAGGTAAGAGGAGGAGTGAACAGTGAACCTCCGATTA 724
760 AACGACCAAGCTCTTCTGAAAGGTAAGAGGAGGAGTGAACAGTGAACCTCCGATTA 819
725 GTGACAGGATGGCTGACCTTGAAGAGCTTCAAGATGCTTCGACAGGAAGCGGAAAC 784

DB 820 GTGACGCGATGACCTGACCTTGAAGAGCTTCAAGATGCTTCCAGAGAGCGGAGAC 879
QY 785 AGGAGCCACTGAGAGGCGCTGAAAGGATCCAGATGCGCTCCGAGAGAGCGCGG 844
DB 880 AGGAGCCACTGAGAGGCGCTGAAAGGATCCAGATGCGCTCCGAGAGAGCGCGG 938
QY 845 AGGAGGTTCCGCGGAGCTGTGGGCTCCGAGGCGAGTCACTGTGAGAGGCGCTCTGC 904
DB 939 AGGAGGTTCCGCGGAGCTGTGGGCTCCGAGGCGAGTCACTGTGAGAGGCGCTCTGC 997
QY 905 CCAAGTAAAGAGAGCGCGGAGCAAGATCAAGGCTGTGCTCCAGGCTGGGAAAGATGACA 964
DB 998 CCAAGTAAAGAGAGCGCGGAGCAAGATCAAGGCTGTGCTCCAGGCTGGGAAAGATGACA 1056
QY 965 GAGAGAGTAAAGTGAAGGAGGAGGCGGCTCAGAGCATCAAGATGAGAGCTGAGACT 1024
DB 1057 RAGAGGTAATGTGAGAGGAGGAGGCGGCTCAGAGCATCAAGATGAGAGCTGAGACT 1116
QY 1025 TGGGCG 1030
DB 1117 CGAGCC 1122

RESULT 10

AL553387 1201 bp mRNA linear EST 31-MAY-2003
LOCUS AL553387 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS0D1075LT03 5-PRIME, mRNA sequence.

ACCESSION AL553387
VERSION AL553387.2 GI:31275201
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
L1,W.B., Gruber,C., Jessee,J. and Polyes,D.
1 (bases 1 to 1201)
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)

COMMENT On Feb 15, 2001 this sequence version replaced gi:12893174.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9799.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1075CP02P1&cluster=9799.r. Contact :

Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0D1075CF02P1.
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1075LT03"
/issue_type="PLACENTA COT 25-NORMALIZED"
/clone_1ib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

FEATURES

source

1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1075LT03"
/issue_type="PLACENTA COT 25-NORMALIZED"
/clone_1ib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 33.1%; Score 863.6; DB 9; Length 1201;
Best Local Similarity 93.5%; Pred. No. 1.6e-203;
Matches 914; Conservative 19; Mismatches 33; Indels 12; Gaps 3;

1 CTCGGTGGCGGTCCGCACTCCGCGGCGTTCGAGGAAATGCTGCGAGACCTTAAGGC 60

```

Db      57 CCGGGATCCGTCGCCCACTCCGGCGGCGTTCGGGGAAATGGCTGAGAGACCTAGAGGC 116
Oy      61 CTGCG-----GAGCTTACTCAACGGGAAACAGCTCTAGATTAATCTGAGTTGTTGAAAT 114
Db      117 CTGCGGCGCTGCGGAACTTCAACGGGAAACAGCTCTAGATTAATCTGAGTTGTTGAAAT 176
Oy      115 ACGAAGCGCTTACTCTGTAACAGTGGCTGACAAACAGTGTGTGTGTGTGTGTGTGTGTGT 174
Db      177 ACGAAGCGCTTACTCTGTAACAGTGGCTGACAAACAGTGTGTGTGTGTGTGTGTGTGTGT 236
Oy      175 GCTTGAACCCAGAGGTTTGTCTGCGCAGGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 234
Db      237 GCTTGAACCCAGAGGTTTGTCTGCGCAGGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 296
Oy      235 AGTGTCCAAAGCTTTTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 294
Db      297 AGTGTCCAAAGCTTTTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 356
Oy      295 GAGGACCTTAAGGCGATCAACCCACAGCAAAAGTTTCAAGGAAACATGGGGTTTTCGA 354
Db      357 GAGGACCTTAAGGCGATCAACCCACAGCAAAAGTTTCAAGGAAACATGGGGTTTTCGA 416
Oy      355 AGGACCACTATGCGCAAGGAGAGGCGCGAGGGAACGCGAGAGCTGACCCACTGAGCCG 414
Db      417 AGGACCACTATGCGCAAGGAGAGGCGCGAGGGAACGCGAGAGCTGACCCACTGAGCCG 476
Oy      415 CCAACCCCAACAGCAGCAGCTGGGCTGTCTCTGCGCGGAGTGGGAGGAGGAGCCCAAGCC 474
Db      477 CCAACCCCAACAGCAGCAGCTGGGCTGTCTCTGCGCGGAGTGGGAGGAGGAGCCCAAGCC 536
Oy      475 ACTGAGCGCGTGGAGCAGTTCCTGACCATTTGCGCGCGCGCGCGAGAGAGCAGTCTT 534
Db      537 ACTGAGCGCGTGGAGCAGTTCCTGACCATTTGCGCGCGCGCGCGAGAGAGCAGTCTT 596
Oy      535 GTCTCCCTGAGAGATTCTGTGAGCCCAAGTCCTGCGCGCGAGAGAGAGAGAGAGAGCC 594
Db      597 GTCTCCCTGAGAGATTCTGTGAGCCCAAGTCCTGCGCGCGAGAGAGAGAGAGAGAGCC 656
Oy      595 TCCGAGGCGCAGTGTGAGAAAGCCCTTCTGAGACCAAGAGCGCCCAAGTCTTCTTCA 654
Db      657 TCCGAGGCGCAGTGTGAGAAAGCCCTTCTGAGACCAAGAGCGCCCAAGTCTTCTTCA 716
Oy      655 GCTGTGAAGAAAGCAGCAGCTCTTCTGAAAGAGTGAAGAGAGAGAGAGAGAGAGAGAG 714
Db      717 GCTGTGAAGAAAGCAGCAGCTCTTCTGAAAGAGTGAAGAGAGAGAGAGAGAGAGAGAG 776
Oy      715 ACCTCGATAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 774
Db      777 ACCTCGATAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 836
Oy      775 AAGCGGGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 834
Db      837 AAGCGGGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 896
Oy      835 AAGCGGGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 894
Db      897 AAGCGGGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 956
Oy      895 GAGCGTCTGCGCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 954
Db      956 GAGCGTCTGCGCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1010
Oy      955 AAGATGACAGAGAGAGT 972
Db      1011 AAGATGACAGAGAGAGT 1028

```

```

RESULT 11
AL550675/c 1018 bp mRNA linear EST 31-MAY-2003
LOCUS AL550675 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS001056YK04 3-PRIME, mRNA sequence.
ACCESSION AL550675

```

```

VERSION AL550675.2 GI:31272492
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE li.W.B., Gruber C., Jessee, J. and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished (2001)
On Feb 15, 2001 this sequence version replaced gi:12887874.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segre@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9799.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS001056BF02NP1&cluster=9799.r. Contact :
Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID: CS001056BF02NP1.

```

FEATURES

source

1..1018

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS001056YK04"

/tissue_type="PLACENTA COT 25-NORMALIZED"

/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with NotI and cloned into the NotI and EcoRV

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match

32.4%; Score 845.2; DB 9; Length 1018;

Best Local Similarity 97.5%; Pred. No. 5.9e-199;

Matches 856; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

```

Oy      1189 AGAATGCGGAAGACTATATCTGCGCAATCTGACCATTTCTGCAAGTGCAGATGAGCT 1248
Db      879 AAGGAATGCGGAGACTATATCTGCGCAATCTGACCATTTCTGCAAGTGCAGATGAGCT 820
Oy      1249 CATTGAGAAACGCGAGATCAAGAGAGCTAAATGAGAGCTGAGAGTCTGATGACCC 1308
Db      819 CATTGAGAAACGCGAGATCAAGAGAGCTAAATGAGAGCTGAGAGTCTGATGACCC 760
Oy      1309 GATTGACAAAGTATGAGAAACATATAGAGAGAGTCTAGGAGAGAGAGAGAGAGAGAGAG 1368
Db      759 GATTGACAAAGTATGAGAAACATATAGAGAGAGTCTAGGAGAGAGAGAGAGAGAGAG 700
Oy      1369 AGAATGAGAAAGCTGCAATTCAGAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1428
Db      699 AGAATGAGAAAGCTGCAATTCAGAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 640
Oy      1429 ATAGAGGCGCTGCTGCTCAAAATGATATGGCCCCCGGTGTGTCAAGTGCAGAGGCC 1488
Db      639 ATAGAGGCGCTGCTGCTCAAAATGATATGGCCCCCGGTGTGTCAAGTGCAGAGGCC 580
Oy      1489 GACTCGGTGTACTGAGATATGATCTGTATCTCAACACGCGGAGAGAGAGAGAGAGAGAG 1548
Db      579 GACTCGGTGTACTGAGATATGATCTGTATCTCAACACGCGGAGAGAGAGAGAGAGAG 520
Oy      1549 CTAAGCTCAGTAAAGAAACAGAGGCTAAAGGCTAAAGAGATGAGTAAAGGCTGAG 1608
Db      519 CTAAGCTCAGTAAAGAAACAGAGGCTAAAGGCTAAAGAGATGAGTAAAGGCTGAG 460
Oy      1609 AAGCCAGCTTCCGAATTCGCGTGTCTCAGCGAGGATTTAAATCTCTTGTGCAAG 1668
Db      459 AAGCCAGCTTCCGAATTCGCGTGTCTCAGCGAGGATTTAAATCTCTTGTGCAAG 400

```

Query Match	32.1%	Score 838.2	DB 13	Length 1050
ORIGIN				
RESULT 12				
EX404270				
LOCUS				
DEFINITION	EX404270 Homo sapiens NEUROBLASTOMA Homo sapiens CDNA clone			
ACCESSION	EX404270			
VERSION	EX404270.1 GI:30764484			
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 1050)			
AUTHORS	Li, W.B., Gruber, C., Jeessee, J. and Polayes, D.			
TITLE	Full-length cDNA libraries and normalization			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 9799.r For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CL0BB022ZG10RPlcluster=9799.r. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CL0BB022ZG10RPl.			
FEATURES				
source	Location/Qualifiers 1..1050 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CL0BB022ZG10" /tissue_type="NEUROBLASTOMA" /clone_lib="Homo sapiens NEUROBLASTOMA" /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."			

	Best Local Similarity	96.7%	Pred No. 3.3e-197	Matches 875	Conservative 12	Mismatches 10	Indels 8	Gaps 3
QY	2	TCGGATGACCCATCCGCCCACTCCGCGCGCTTCCGCGGAAATGCTGCGAGACCTTAAGAGCC	61					
Db	52	TCCCGGGAATTCGCCCCCACTCCGCGCGCTTCCGCGGAAATGCTGCGAGACCTTAAGAGCC	111					
QY	62	TGCG-----GAGCTTAAGTCAATGCGGAAAGAGCTCTAAGATAATCTGATGTTGAAATA	115					
Db	112	TGCGGCTCTGCGASATTACTCAAGGAAAGAGCTCTAAGATAATCTGATGTTGAAATA	171					
QY	116	GGAAGCTGTAACTCCGGAACAGTGGCTGAACAAGTGTGTGTGAGCTGCTGTCTG	175					
Db	172	GGAAGCTGTAACTCTGTGAACAGTGGCTGAACAAGTGTGTGTGAGCTGCTGTCTG	231					
QY	176	CTTGGACCCAGAGGTTTGTCTGSCAGAGGTTTTTGTTATTTAGATTTCAAGGAAA	235					
Db	232	CTTGGACCCAGAGGTTTGTCTGSCAGAGGTTTTTGTTATTTAGATTTCAAGGAAA	291					
QY	236	GTGTTCMAAGCTTTCAAGTGTGGAGCAGGTATGACCAAAAGCGACCCGAGCAATGAG	295					
Db	292	GTGTTCMAAGCTTTCAAGTGTGGAGCAGGTATGACCAAAAGCGACCCGAGCAATGAG	351					
QY	296	AGGCACTTAAGGCCATTAACCAACAGACAAAGTTCAGAGAAAACAATGGGTTTTTCGAA	355					
Db	352	AGGCACTTAAGGCCATTAACCAACAGACAAAGTTCAGAGAAAACAATGGGTTTTTCGAA	411					
QY	356	GGAACCACTATCCGACAGAGAGAGGCGCAGGAGACGCGAGAGCTGACCCCACTGAGCCGC	415					
Db	412	GGAACCACTATCCGACAGAGAGAGGCGCAGGAGACGCGAGAGCTGACCCCACTGAGCCGC	471					
QY	416	CACCCCAACAGCAGCAGCTGGGCTGTCCCTCGCGGCAATGGAGGCAACCCCAAGGCA	475					
Db	472	CACCCCAACAGCAGCAGCTGGGCTGTCCCTCGCGGCAATGGAGGCAACCCCAAGGCA	531					
QY	476	CTGAGCGCGTGGAGCAGTTCTCTGACCAATTCGCGCGCGCGCGGAGAGAGCATGCTG	535					
Db	532	CTGAGCGCGTGGAGCAGTTCTCTGACCAATTCGCGCGCGCGCGGAGAGAGCATGCTG	591					
QY	536	TCTCCCTGAGAGATTCTGTGTAAGCCCAAGTCTTGCCCGCAACAGACGCGGAGCAGCTT	595					
Db	592	TCTCCCTGAGAGATTCTGTGTAAGCCCAAGTCTTGCCCGCAACAGACGCGGAGCAGCTT	651					
QY	596	CCGAGGCGAGCGTGGAAAGCGCTTCTGAACACAGACAGCGGCCCTCACTGTCTTCCACAG	655					
Db	652	CCGAGGCGAGCGTGGAAAGCGCTTCTGAACACAGACAGCGGCCCTCACTGTGTCCACAG	711					
QY	656	CTGAGGAAGGAACGACAGCGCTCTTCTGAAAAGGTGAAGAGAGGGGATGACCAAGATGACA	715					
Db	712	CTGTTAAGGAACGACAGCGCTCTTCTGAAAAGGTGAAGAGAGGGGATGACCAAGATGACA	771					
QY	716	CCTCCGATGATGACAGAGATGAGCTGACCTTGAAGAAGCTTCAGAAATCGCCTTGCAGGA	775					
Db	772	CCTCCGATGATGACAGAGATGAGCTGACCTTGAAGAAGCTTCAGAAATCGCCTTGCAGGA	831					
QY	776	AGCGGGAACAGAGAGCCCACTGAAGAGCGCCCTGAAAGAGATTCAGAGTCGCTTGGGGAAGA	835					
Db	832	AGCGGGAACAGAGAGCCCACTGAAGAGCGCCCTTGAAGAAGCTTCAGAGTCGCTTGGGGAAGA	890					
QY	836	AGCGCGGGAAGAGAGGTTCGCGCGGAGACTGTGGGCTCCAGAGCCAGTGAACACTGTGAGAG	895					
Db	891	AGCGCGGGAAGAGAGGTTCGCGCGGAGACTGTGGGCTCCAGAGCCAGTGAACACTGTGAGAG	949					
QY	896	GCCTC 900						
Db	950	GCCTC 954						
RESULT 13								
LOCUS	CD364822	850 bp	mRNA	linear	EST 29-MAY-2003			
DEFINITION	UI-H-PT2-bjm-d-20-0-UI.s1 NCI CGAP PT2 Hom sapiens cDNA clone							
	UI-H-PT2-bjm-d-20-0-UI 3', mRNA sequence.							

ACCESSION CS064822
 VERSION CS064822.1 GI:31148912
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 850)
 AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/nciccap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: Dr. Gary W. Hunninghake, U of I
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/cgap.html>
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES
 source
 location/Qualifiers
 1..850
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-H-FT2-bjm-d-20-0-UI"
 /issue_type="Alveolar Macrophage"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_id="NCI CGAP FT2"
 /note="Organ: Lung; Vector: pRTT3-Pac (Pharmacia) with a
 modified polylinker; Site_1: Ecor I; Site_2: Not I;
 NCI CGAP FT2 is a subcloned cDNA library constructed from
 a pool of 81 RNA samples from Alveolar Macrophages
 challenged with different treatments. The library was
 subcloned according to Bonaldi, Lennon and Soares, Genome
 Research, 6:791-806, 1996. The tissue was provided by Dr.
 Gary W. Hunninghake of the University of Iowa.
 TAG LIB=UI-H-FT2
 TAG TISSUE=Human Lung Alveolar Macrophage
 TAG_SEQ=GCCCATGCCG"

ORIGIN
 Query Match 31.8%; Score 830.4; DB 14; Length 850;
 Best Local Similarity 99.6%; Pred. No. 2.7e-195;
 Matches 831; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1258 ACCGCAATCGACGAGAGCTTAATGAGACCTGAGAGTGTGAGGACCGATTTGACA 1317
 Db 850 ACGGCAATCGACGAGAGCTTAATGAGAGCTGAGAGTGTGAGGACCGATTTGACA 791

1318 AGTATAGAAACAATGAGAGCTGAGAGGACCAAGGATTAAGGATTTGAG 1377
 Db 790 AGTATAGAAACAATGAGAGCTGAGAGGACCAAGGATTAAGGATTTGAG 731

1378 AAAGCTGCAATTCAGAGTGGCAAGAACTCAAGATCTTCCAGCTGTGATAGAGCG 1437
 Db 730 AAAGCTGCAATTCAGAGTGGCAAGAACTCAAGATCTTCCAGCTGTGATAGAGCG 671

1438 CCTGTGCTCAAAATGATTTGGCCGGGCTTTCAGCTGAGCGGACCGGACCTCGTG 1497
 Db 670 CCTGTGCTCAAAATGATTTGGCCGGGCTTTCAGCTGAGCGGACCGGACCTCGTG 611

1498 TACTGCAATTAATGATCTGATCTCAACAACGCGGACGACCAATGAAGTTTAAAGTCA 1557
 Db 610 TACTGCAATTAATGATCTGATCTCAACAACGCGGACGACCAATGAAGTTTAAAGTCA 551

1558 GGTAAAGAAAGAGAGCAAGGCTTAAGAAAGATGAAGATGAGGACGAGAGAGCCAGT 1617
 Db 550 GGTAAAGAAAGAGAGCAAGGCTTAAGAAAGATGAAGATGAGGAGAGAGCCAGT 491

1618 CTTCCGAATTCGAGTCTCAGGACAGTATTAATCTTCTGTGACAGAGACCACT 1677
 Db 490 CTTCCGAATTCGAGTCTCAGGACAGTATTAATCTTCTGTGACAGAGACCACT 431

1678 CCGAAGAAAGAGAGCAAGTGAAGAGGACAGTGTGCTTCCGCGAGTGAAGCA 1737
 Db 430 CCGAAGAAAGAGAGCAAGTGAAGAGGACAGTGTGCTTCCGCGAGTGAAGCA 371

1738 CTGGGAGAGAGAGCTTGTGAAGAGACGACCGCTGTGGGAGAGGATCAATTTAC 1797
 Db 370 CTGGGAGAGAGAGCTTGTGAAGAGACGACCGCTGTGGGAGAGGATCAATTTAC 311

1798 AATGAGTAAAGCCAGAAAGACCTGCTCTCCGCGTCACTGTTATTAATGATG 1857
 Db 310 AATGAGTAAAGCCAGAAAGACCTGCTCTCCGCGTCACTGTTATTAATGATG 251

1858 TATCACTTGAAGGTTGGCTTCTGAGACCCCTTCCGTTCTTTTGATAGCCATCCCTGG 1917
 Db 250 TATCACTTGAAGGTTGGCTTCTGAGACCCCTTCCGTTCTTTTGATAGCCATCCCTGG 191

1918 GCGTGTCCAGAGCTTGGAGTTGACAGCTTGTGTTAAGCTGATCAGACACCGGCTGAC 1977
 Db 190 GCGTGTCCAGAGCTTGGAGTTGACAGCTTGTGTTAAGCTGATCAGACACCGGCTGAC 131

1978 CATCAGCGGAGAGAGAGCCCATGTCAGATGCTCTCTGCTGCGCTGTTCATTCCTTA 2037
 Db 130 CATCAGCGGAGAGAGAGCCCATGTCAGATGCTCTCTGCTGCGCTGTTCATTCCTTA 71

2038 GTCTGTCCAGAGCTTCTGTCTGACCTGTTTCCAAAGCTGTAACTTACTGTGTAA 2091
 Db 70 GTCTGTCCAGAGCTTCTGTCTGACCTGTTTCCAAAGCTGTAACTTACTGTGTAA 17

RESULT 14
 AK014422 1614 bp mRNA linear HTC 20-SEP-2003
 LOCUS AK014422
 DEFINITION Mus musculus 18 days pregnant adult female placenta and extra
 embryonic tissue cDNA, RIKEN full-length enriched library,
 clone:3830408B01 product:death inducer-obliterator-1, full insert
 sequence.

ACCESSION AK014422
 VERSION AK014422.1 GI:12852263
 KEYWORDS HTC; CAP trapper
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

REFERENCE 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20493374
 PUBMED 11042159

REFERENCE 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Kono, H., Akiyama, J., Nishi, K., Kitamura, T., Tashiro, H., Itoh, M.,
 Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Oogawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicephillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913

PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6
AUTHORS (bases 1 to 1614)
Adechi, J., Alizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukushima, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Iwaw, M., Kanakawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koyama, K., Kunitzaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Momura, K., Numazaki, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Shukri, H., Tagami, M., Tagawa, A., Takahashi, P., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Yamatsuta, M., and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Shohiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-research.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3']. cDNA was prepared by using trihalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 50.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified Bluescript KS(+) after bulk excision from Lambda PUC1. Cloning sites, 5' end: SalI, 3' end: BamHI. Host: DH10B.
FEATURES
SOURCE Location/Qualifiers
1..1614
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM:DB:3830408B01"
/db_xref="MGI:1910351"
/db_xref="taxon:10090"
/clone="3830408B01"
/sex="Female"
/tissue_type="placenta and extra embryonic tissue"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="18 days pregnant adult"
331..1614
/note="deach inducer-oblierator-1 (MGD|MGI:1344352, GBI|W238332, evidence: BLASTN, 99%, match=2599) putative"
misc_feature
Query Match 31.3%; Score 817.8; DB 11; Length 1614;
Base Local Similarity 73.1%; Pred. No. 4.4e-192;
Matches 1127; Conservative 0; Mismatches 387; Indels 27; Gaps 5;
ORIGIN

QY 21 TCCGCGGCGGTTGCGGAAATGCGTAGAGACCTGCGGAGCTTACTCCACG 80
DB |||||
QY 100 TCCGTGTGAGCTCTGAAATGCGTGGGATCCCGCGGCCGGGAGCTTTTAAGAG 159
DB |||||
QY 81 GAACAGCCTCTAGATATCTGAGTTGTGAAATACGAAGCCTGTACTGTGAAACAGTG 140
DB |||||
QY 160 CAGTCCCACTGTCCCTTGTGTGTGAAAGCTCCGGAATCTTCTCATGTATGAACTGTG 219
DB |||||
QY 141 GCTGACAACTGTGTGTGTGAGCCTGCGTGTCTGTGTGAAACCAAGGTTCTGTCC 200
DB |||||
QY 220 ACTGACAACTGAGGAGTGAG--GCTTGGCCGTCTGCTTGAAC-----TGCGCC 265
DB |||||
QY 201 AGGTTTTGTGTGTATTTAGATTT--CAGGAAAGTGTCCAGCTTCAAGTTGTGAG 259
DB |||||
QY 266 CAGGCTTAAATTTATGTAGGATTTCCAGCAAGGTTTCAAGTTTCAAGTTTGGGA 325
DB |||||
QY 260 CAGTATGAGACGAAAGCGACCCGAGCAATGAGAGGCACTTAAGGCATCAACCCA 319
DB |||||
QY 326 CAGGTATGTATTAAGGCACTTGAACATGAGGAGCACCAGCTATCAAAACCA 385
DB |||||
QY 320 CCAGCAAGAGTTCAAGAAACATGCGGTTTTCGAAAGCACTATGCGCAAGCGAGG 379
DB |||||
QY 386 CAGTATGAGAGTTCAAGAAACCTGGGTTTTCGAAAGCAAGATTCGCAACGTAAG 445
DB |||||
QY 380 GCGGAGGAGACCGGAGGCTGACCCACTGGAACCCCAACCCACAGCAGCTGGGC 439
DB |||||
QY 446 GTGCGAGAGACGAGGAGTGACCCCACTGAGCAGCAAC-----ACAGCAGCATPACC 439
DB |||||
QY 440 TGTCCCTGCGCGAGTGTGAGGAGCGACCCAGCACTGAGCGCTGTGAGAGCTTCTGA 499
DB |||||
QY 500 TTTCTCTGCGCGAGTGTGAGGAGCGGCAACCAACGTAAGAGGTTCTTGA 559
DB |||||
QY 500 CCATTGCGCGCGCGCGAGGAGAGAGCATGCTGTCTTCTGTGAGATTTCTGTGAGC 559
DB |||||
QY 560 CCAAGTTCTGCGCGCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 619
DB |||||
QY 620 CCACATCTTCCACAGTCACTGATGTGAGACAGCTTCCAGGAGGAGAGAGAGAGAG 679
DB |||||
QY 620 CTGAGACGAGAGGCGCCCGAGTGTGCTTCCAGAGCTGTGAGAGAGAGAGAGAG 679
DB |||||
QY 680 CTGATATCAAGATGCTCTGTA-----TGTATCTCTTGAAGAGAGATCTGCTCTT 733
DB |||||
QY 734 CTGAAAGGCTGAAAG 739
DB |||||
QY 740 TGAATTTGAAGAGCTTCAAGATGCGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAG 799
DB |||||
QY 794 TTAAGTTGAGGAGATTTCAAGAACCGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAG 853
DB |||||
QY 800 GCGCCCTGAAAGAGAGATTCAGAGTGTGCGTGTGAGAGAGAGAGAGAGAGAGAGAG 859
DB |||||
QY 854 GCTTCTTGAAG 913
DB |||||
QY 860 AGACTGTGCGCTCCAGAGCAGTGTGACACTGTGTGAGAGAGAGAGAGAGAGAGAGAG 919
DB |||||
QY 914 AATCTGTGAGAGTGTCAATATGAGCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 973
DB |||||
QY 920 CCGAGAGAGATCAAG 979
DB |||||
QY 974 CTGAGAGTGTGAG 1033
DB |||||
QY 980 AGGGAAG 1039
DB |||||
QY 1034 AAGGGAAG 1093
DB |||||
QY 1040 CTGATGTGAG 1099
DB |||||
QY 1094 CTGATGTGAG 1153
DB |||||
QY 1100 ACAAGTTATGATTTGTGCTGTGACCCCTGTGAAAGATGTTTCAAGAGAGAGAGAGAG 1159
DB |||||

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 23, 2004, 07:35:03 ; Search time 192.11 Seconds
(without alignments)
8281.924 Million cell updates/sec

Title: US-09-787-016A-2

Perfect score: 1 tccgtgtagctctggaat.....tttcttctctgtgctg99 2867

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Database: Issued Patents, NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/Backfillseq.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	77.8	2.7	7218	1	US-08-232-463-14
C 2	55.4	1.9	6256	4	US-09-620-312D-269
C 3	40.2	1.4	7218	1	US-08-232-463-14
C 4	37.8	1.3	505	4	US-09-621-976-15639
C 5	37.8	1.3	176373	3	US-09-128-155-17
C 6	37.2	1.3	6755	3	US-08-931-999-4
C 7	36.6	1.3	248	3	US-09-007-005-32
C 8	36.6	1.3	248	3	US-09-244-796-32
C 9	36.6	1.3	277	3	US-09-007-005-3
C 10	36.6	1.3	277	3	US-09-244-796-3
C 11	36.4	1.3	765	4	US-09-328-352-3047
C 12	36.4	1.3	474	4	US-09-621-976-18033
C 13	35.8	1.2	1492	4	US-09-196-520-7
C 14	35.6	1.2	289	3	US-09-007-005-17
C 15	35.6	1.2	289	3	US-09-244-796-17
C 16	35.6	1.2	2244	3	US-09-061-764A-18
C 17	35.6	1.2	2824	4	US-09-016-434-1349
C 18	35.6	1.2	190	4	US-09-188-930-57
C 19	35.6	1.2	190	4	US-09-312-283C-57
C 20	35.6	1.2	190	4	US-09-312-283C-402
C 21	35.6	1.2	472	4	US-09-621-976-18485
C 22	34.8	1.2	1230025	4	US-09-198-452A-1
C 23	34.6	1.2	245	1	US-08-139-937-8
C 24	34.6	1.2	349	5	PCT-US93-11310-8
C 25	34.4	1.2	295	4	US-09-621-976-8976
C 26	34.2	1.2	1501	3	US-09-196-520-1
C 27	34	1.2	732	4	US-09-134-000C-348

C 28	34	1.2	1146	1	US-08-487-810-1	Sequence 1, Appl
C 29	33.8	1.2	505	4	US-09-621-976-15639	Sequence 15639, A
C 30	33.6	1.2	424	4	US-09-621-976-7975	Sequence 7975, Ap
C 31	33.6	1.2	1886	4	US-09-594-506-31	Sequence 31, Appl
C 32	33.6	1.2	2278	4	US-09-620-312D-1003	Sequence 1003, Ap
C 33	33.6	1.2	2278	4	US-09-148-545-78	Sequence 78, Appl
C 34	33.6	1.2	2278	4	US-09-148-545-131	Sequence 131, App
C 35	33.4	1.2	1384	4	US-09-205-258-225	Sequence 225, App
C 36	33.4	1.2	5066	4	US-08-961-527-79	Sequence 79, Appl
C 37	33.4	1.2	5345	1	US-08-044-618-7	Sequence 7, Appl
C 38	33.4	1.2	11049	4	US-10-204-708-23	Sequence 23, Appl
C 39	33.4	1.2	11384	4	US-08-961-527-45	Sequence 45, Appl
C 40	33.4	1.2	202001	4	US-09-734-674-3	Sequence 3, Appl
C 41	33.2	1.2	1497	4	US-09-489-039A-6080	Sequence 6080, Ap
C 42	33.2	1.2	1950	4	US-09-614-891-2	Sequence 2, Appl
C 43	33.2	1.2	2135	4	US-08-933-711B-17	Sequence 17, Appl
C 44	33.2	1.2	2501	4	US-09-614-891-3	Sequence 3, Appl
C 45	33	1.2	364	4	US-09-621-976-17202	Sequence 17202, A

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHRIFFLINGER, F.
APPLICANT: FALKNER, P. G. F.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: BP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-8109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9DC-F18
US-08-232-463-14
Query Match 2.7%; Score 77.8; DB 1; Length 7218;

[illegible]

```

RESULT 2
US-09-620-312D-269
: Sequence 269, Application US/09620312D
: Patent No. 6569662
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Liu, Chenghua
: APPLICANT: Asundi, Vinod
: APPLICANT: Zhang, Jie
: APPLICANT: Ren, Feiyen
: APPLICANT: Chen, Rui-hong
: APPLICANT: Zhao, Qing A.
: APPLICANT: Wehrman, Tom
: APPLICANT: Xue, Aidong J.
: APPLICANT: Yang, Yonghong
: APPLICANT: Wang, Jian-Rui
: APPLICANT: Zhou, Ping
: APPLICANT: Ma, Yundong
: APPLICANT: Wang, Dunrui
: APPLICANT: Wang, Zhiwei
: APPLICANT: John Tillinghaaf
: APPLICANT: Dmanan, Radoje T.
: TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and
: FILE REFERENCE: 784CIP28
: CURRENT APPLICATION NUMBER: US/09/620,312D
: CURRENT FILING DATE: 2000-07-19
: PRIOR APPLICATION NUMBER: 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: 09/468,725
: PRIOR FILING DATE: 2000-01-21
: NUMBER OF SEQ ID NOS: 1105
: SOFTWARE: pc_fl_genes Version 1.0
: SEQ ID NO 269
: LENGTH: 6256
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (257)..(5443)
: US-09-620-312D-269

```

Query Match	1.9%	Score 55.4	DB 4	Length 6256
Best Local Similarity	59.0%	Pred. No. 4	78-06	
Matches	95	Conservative	0	Mismatches 66
			Indels	0
			Gaps	0

Qy	1018	CCCAATGCCCTGTACTGACATCTGCGCCAGCCTTCACAACAACAGCTTTATGATCTCTGT	1077
Db	1469	CCGAGCAAGCAGATGTGGGTTTGGCAAAAACCAATGCGCAACAGTTTATGTTGGCTGT	1528
Qy	1078	GATGGCTGTGAGAGATGTTCCACGCTGACGTGTGTGGTATTTCTGAGCGCCGAGGGCGG	1137
Db	1529	GCGAGATGTATGTGCTGTTCATGTGATGTGTGTGGTTAAGTCTTTCTCAAGCACAG	1588
Qy	1138	CTCTCGAAGAAGACCGGGAGAATCAATCATCTGCCCAATTG	1178
Db	1589	CAGATGGCCGAGGAAGACCAAGATATGTCTGTGTAAAAATG	1629

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```



```

RESULT 7
US-09-007-005-32
Sequence 32, Application US/09007005B
Patent No. 6258558
GENERAL INFORMATION:
APPLICANT: Szostak, Jack W.
APPLICANT: Szostak, Robert, Richard W.
APPLICANT: Liu, Rihne
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
TITLE OF INVENTION: FUSIONS
FILE REFERENCE: 00786/350003
CURRENT APPLICATION NUMBER: US/09/007.005B
CURRENT FILING DATE: 1998-01-14
EARLIER APPLICATION NUMBER: 60/035,963
EARLIER FILING DATE: 1997-01-27
EARLIER APPLICATION NUMBER: 60/064,491
EARLIER FILING DATE: 1997-11-06
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSBQ for Windows Version 4.0
SEQ ID NO 32
LENGTH: 248
TYPE: RNA
ORGANISM: Homo sapiens
US-09-007-005-32

```

```

: Sequence 32, Application US/09244796
: Patent No. 6281344
:
: GENERAL INFORMATION:
: APPLICANT: Szostak, Jack W.
: APPLICANT: Szostak, Roberts, Richard W.
: APPLICANT: Liu, Rihne
:
: TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
: TITLE OF INVENTION: FISHONS
:
: FILE REFERENCE: 00786/350007
:
: CURRENT APPLICATION NUMBER: US/09/244,796
:
: CURRENT FILING DATE: 1999-02-05
:
: EARLIER APPLICATION NUMBER: 60/035,963
:
: EARLIER FILING DATE: 1997-01-27
:
: EARLIER APPLICATION NUMBER: 60/064,491
:
: EARLIER FILING DATE: 1997-11-06
:
: EARLIER APPLICATION NUMBER: 09/007,005
:
: EARLIER FILING DATE: 1998-01-14
:
: NUMBER OF SEQ ID NOS: 33
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO 32
:
: LENGTH: 248
:
: TYPE: RNA
:
: ORGANISM: Homo sapiens
:
: US-09-244-796-32

```

```

QY      768  GGGCAGTCAAGATGCCCTGAGGAAGAGCGCAGAGGAAAGATTCTGCCAACTGGAG  827
Db      1  KGGGGRARCRAPARUPURCRURCRUPURUPARAPARUPURARCRARURGRRC  60

QY      828  TGTCCAAATTAGCAGTGCAGACGAGCAGACCTCTCTGTAAAGCAGAGCTGAGCTAG  887
Db      61  RURGAPARARGABARCRARGABARARCRURBARURCRURCRABARABARABARA  120

QY      888  TCAGGACCAAGTCTCCAGTCAAGACAGATGACATAGAAATCACTTGGAGAGGAGGC  947
Db      121  RCRURRGRURGRCRGRURAPARARCRGRURCRURGRARARCRARCRCRURGARA  180

QY      948  GACTCAGGAAATATACAGAGAAACCCAGGAGGAGCGCAACCAAGCTGAGTGTG  1006
Db      181  RARCRARARARARCRURGRAPARCARAGCRURGRCRGURUPARARURURCRUR  239

```

RESULT 9
US-09-007-005-3
; Sequence 3, Application US/09007005B
; Patent No. 6258558

Query Match	1.2%	Score 35.8	DB 3	Length 1492
Best Local Similarity	46.9%	Pred. No. 2,2		
Matches 112	Conservative	0	Mismatches 127	Indels 0
			Gaps	0

QY	1381	CCTGTCTGAGAGGCTCTCTGTCCTCTTAATGATGTGGCCCTGGTGTTCACGTGTACGA	1440
Db	786	CTTGAGGAAAAAGTCGCTCTGTTTTCCTATGATATATAGCTTATGTATGACGAGAACCT	727
QY	1441	CAGCTTGACTCTGTATTTACGTATATGATCTGCATTTTCCAAACAGCAGAGCTTACATG	1500
Db	726	CTCCCTGCAATGGCTTTGGCATTAATGAAAGCATACCTCTGCACCTTCAAACCTTGCTGG	667
QY	1501	AGATTTCTAAAGTTCAGGTAAAGAACAAAAACCAACCAAGAAAGGTCAAGACGAAG	1560
Db	666	GCGGGTATATATTTTCAAACCTTCCACAAACCCCTCAACACTGATGTCAAGGCACT	607
QY	1561	CCAGAAAAGTTCAGTCTTCCAAATGACAGTGTTCAGTGGGGGATTAANAATCTCTTCTGT	1619
Db	606	ATACTTCTCTCAGATATTTTCAACGAGATTGTCAAGGTTCAATCATATATCTAGTCT	548

RESULT 14
US-09-007-005-17
Sequence 17, Application US/09007005B

```

? Patent No. 6258558
? GENERAL INFORMATION:
? APPLICANT: Szostek, Jack W.
? APPLICANT: Roberts, Richard W.
? APPLICANT: Liu, Rihc
? TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
? TITLE OF INVENTION: FUSIONS
? FILE REFERENCE: 00786/35003
? CURRENT APPLICATION NUMBER: US/09/007,005B
? CURRENT FILING DATE: 1998-01-14
? EARLIER APPLICATION NUMBER: 60/035,963
? EARLIER FILING DATE: 1997-01-27
? EARLIER APPLICATION NUMBER: 60/064,491
? EARLIER FILING DATE: 1997-11-06
? NUMBER OF SEQ ID NOS: 33
? SOFTWARE: fastseq for Windows Version 4.0
? SEQ ID NO 17
? LENGTH: 289
? TYPE: RNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Translation template
? NAME/KEY: misc feature
? LOCATION: (1) -- (289)
? OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

```

	Query Match	1.2%;	Score 35.6;	DB 3;	Length 289;	
	Best Local Similarity	9.0%;	Pred. No. 0.86;			
	Matches	24;	Conservative 104;	Mismatches 140;	Indels 0;	Gaps 0;
Dy	728	GGAGAAAGCGAGACAGAAGAACTGTGGAAGAGTCCCTGAGAGCAGTCAGAAATCCCTGA	787			
	: : : : : : : : : : : : : : :					
Dd	2	GRGGRARCABARURBARCRUBARURURBARCRABARURURBARCRARURGRNRNR	61			
Oy	788	GGAAGAACGCAGAGAGAGAAAGATTCTGCCGAACCTGGAGATGTCCAAATAGCCAGTCCG	847			
	: : : : : : : : : : : : : : : : : : :					
Dd	62	SRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNR	121			
Oy	848	AGCAGACACAACCTCTCTGTAAGAGAGAGCTCTAGAGCTACTCAGAGACCAGTCCCACT	907			
	: : : : : : : : : : : : : : : : : : :					
Dd	122	SRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNR	181			
Oy	908	CAGAGACAGATGACATGAAAATCAGTTTGGAAGGAAAGCGAATCAGAGAAATACAGAG	967			
	: : : : : : : : : : : : : : : : : : :					
Dd	182	SRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNRSRNNR	241			
Oy	968	AAAAACCCAGGGAAGCGGCAAAACCA	995			
	: : : : : : : : : : : :					
Dd	242	CRUCRCRURRGCRGRCRUAAAAAAAAA	269			

RESULT 15-796-17
 US-09-244-796-17
 : Sequence 17, Application US/09244796
 : Patent No. 6281344
 : GENERAL INFORMATION:
 : APPLICANT: Szostak, Jack W.
 : APPLICANT: Roberts, Richard W.
 : APPLICANT: Liu, Rhee
 : TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
 : TITLE OR INVENTION: FUSIONS
 : FILE REFERENCE: 00786/350007
 : CURRENT APPLICATION NUMBER: US/09/244,796
 : CURRENT FILING DATE: 1999-02-05
 : EARLIER APPLICATION NUMBER: 60/035,963
 : EARLIER FILING DATE: 1997-01-27
 : EARLIER APPLICATION NUMBER: 60/064,491
 : EARLIER FILING DATE: 1997-11-06
 : EARLIER APPLICATION NUMBER: 09/007,005
 : EARLIER FILING DATE: 1998-01-14
 : NUMBER OF SEQ ID NOS: 33
 :

```

SOFTWARE: Patsesq for Windows Version 4.0
SEQ_ID NO 17
LENGTH: 289
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Translation template
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)_(1289)
OTHER INFORMATION: n = A,T,C or G
--S-09-244-796-17

```

OTHER INFORMATION: n = A, T, C or G
US-09-244-796-17

Query Match	1.2%	Score 35.6	DB 3	Length 289
Query Match	1.2%	Score 35.6	DB 3	Length 289

Best Local Similarity: 9.0%; Pred. No. 0.86;
Matches 24; Conservative 104; Mismatches 140; Indels 0; Gaps 0;

728 GGAGAAAGCGAGAGCAAGAACTGTGAGAGGTCCCTGAGAGGCAGTCAGAAATCGCCTGA 787

Db 2 GGRGRACRARURUTRACRURARURUTRARCRARURUTRACCRARURGRNNT 61

788 GGAAGAGCGCAGAGGAAGATTCTGCCGAACCTGGAGTGTCCAAATAGGCAGTGCCG 847

[illegible]

848 AGCAGGACGACCTCTCTGTAGCAGGAGCCTGAGGCTAGTCAGGGACCAAGTGTCCAGT 907

[illegible]

908 CAGAGCAGATGACATAGAAATCAGTTGGAAGGAGCGACTCAGGGAATACAGAGG 967

Db **182** SRNRNRSRNRRSRPNRNSRNRRNSRCPAGRCRURGRCGRURARAR **241**

968 AAAACCCGAGGAGCGCAACCAA 995

Db 242 CRURCRURGRCRGRUAAAAAAAA 269

Search completed: April 23, 2004, 18:09:06
Job time : 197.11 secs

Job time : 197.11 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 14:17:01 / Search time 1363.62 Seconds
(Without alignments)
9479.366 Million cell updates/sec

Title: US-09-787-016a-2

Perfect score: 2867

Sequence: 1 tccgtgtagctcgtgaat.....cttgcctctgctgctg 2867

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	558	19.5	1764	9	US-09-925-301-327 Sequence 327, App
2	354.6	12.4	645	15	US-10-029-386-25091 Sequence 25091, A
3	288	10.0	478	10	US-09-918-995-19127 Sequence 19127, A
4	117.8	4.1	536	15	US-10-029-386-11381 Sequence 11381, A
5	65	2.3	65	10	US-09-908-975-25239 Sequence 25239, A
6	57.8	2.0	915	9	US-09-764-864-471 Sequence 471, App
7	57.8	2.0	2320	15	US-10-175-523-112 Sequence 112, App
8	55.4	1.9	5975	10	US-09-873-357C-112 Sequence 112, App
9	55.4	1.9	6256	15	US-10-037-270-269 Sequence 269, App
10	55.4	1.9	6256	16	US-10-117-722-269 Sequence 269, App
11	55.4	1.9	6948	10	US-09-930-213-315 Sequence 315, App
12	50.4	1.8	653	15	US-10-184-644-402 Sequence 402, App
13	50.4	1.8	653	15	US-10-184-634-402 Sequence 402, App
14	48	1.7	469	13	US-10-085-783A-56748 Sequence 56748, A

15	48	1.7	469	16	US-10-242-535A-56748	Sequence 56748, A
16	48	1.7	483	9	US-09-864-761-2140	Sequence 2140, App
17	47.4	1.7	594	13	US-10-142-426-10	Sequence 10, App1
18	47.4	1.7	594	15	US-10-123-155-10	Sequence 10, App1
19	47.4	1.7	594	15	US-10-146-731-10	Sequence 10, App1
20	47.4	1.7	594	15	US-10-140-472-10	Sequence 10, App1
21	47.4	1.7	594	15	US-10-141-761-10	Sequence 10, App1
22	47.4	1.7	594	15	US-10-142-885-10	Sequence 10, App1
23	47.4	1.7	594	15	US-10-158-790-10	Sequence 10, App1
24	47.4	1.7	594	16	US-10-137-871-10	Sequence 10, App1
25	47.4	1.7	594	16	US-10-140-923-10	Sequence 10, App1
26	47.4	1.7	594	16	US-10-141-756-10	Sequence 10, App1
27	47.4	1.7	594	16	US-10-141-759-10	Sequence 10, App1
28	47.4	1.7	594	16	US-10-140-805-10	Sequence 10, App1
29	47.4	1.7	594	16	US-10-140-864-10	Sequence 10, App1
30	42.8	1.5	114615	13	US-10-087-192-676	Sequence 676, App
31	41.4	1.4	745	13	US-10-142-426-68	Sequence 68, App1
32	41.4	1.4	745	15	US-10-123-155-68	Sequence 68, App1
33	41.4	1.4	745	15	US-10-146-731-68	Sequence 68, App1
34	41.4	1.4	745	15	US-10-140-472-68	Sequence 68, App1
35	41.4	1.4	745	15	US-10-141-761-68	Sequence 68, App1
36	41.4	1.4	745	15	US-10-142-885-68	Sequence 68, App1
37	41.4	1.4	745	15	US-10-158-790-68	Sequence 68, App1
38	41.4	1.4	745	16	US-10-137-871-68	Sequence 68, App1
39	41.4	1.4	745	16	US-10-140-923-68	Sequence 68, App1
40	41.4	1.4	745	16	US-10-141-756-68	Sequence 68, App1
41	41.4	1.4	745	16	US-10-141-759-68	Sequence 68, App1
42	41.4	1.4	745	16	US-10-140-805-68	Sequence 68, App1
43	41.4	1.4	745	16	US-10-140-864-68	Sequence 68, App1
44	40.8	1.4	255	13	US-10-085-783A-34088	Sequence 34088, A
45	40.8	1.4	255	16	US-10-242-535A-34088	Sequence 34088, A

ALIGNMENTS

RESULT 1
US-09-925-301-327
Sequence 327, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05892
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 327
LENGTH: 1764
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1398)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1758)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1759)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1762)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-327
Query Match 19.5%; Score 558; DB 9; Length 1764;
Best Local Similarity 74.2%; Pred. No. 3,4e-166;

OY 821 CTGGAG 827
DB 45 CTGTGG 39

RESULT 3

US-09-918-995-19127
Sequence 19127, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT FILING DATE: 2001-07-30
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 19127
LENGTH: 478
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(478)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-19127

Query Match 10.0%; Score 288; DB 10; Length 478;
Best Local Similarity 77.7%; Pred. No. 2e-80;
Matches 348; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

OY 850 CAGACAGACCTCTCTGTAAAGAGAGCTGAGCTGAGGACCACTGCTCCAGTCA 909
DB 31 CGGACAGACCTCTCTGTAAAGAGAGCTGAGCTGAGGACCACTGCTCCAGTCA 90
OY 910 GAGACAGATGATGAAATGCAATGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 969
DB 91 GGGAAAGATGAG 150
OY 970 AACCCAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1029
DB 151 GAGCTGAGAGAGCTGTGGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 210
OY 1030 TACTGATCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1089
DB 211 TATTGATTTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 270
OY 1090 GAGTGTTCATGAGTGAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1149
DB 271 GATGATTTTCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 330
OY 1150 AACGGGAGAGATACATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1209
DB 331 AATGGGAGAGATACATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 390
OY 1210 GGTAGGCGACCAATGAGAGAGCTGTGGGTGAGATCTGTGGGTGAGTGTGAGTGT 1269
DB 391 TCAGAAACCGCAGATGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 450
OY 1270 TGCAGAGATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1297
DB 451 TGTACAGATATAGAAACAATAGAGCAGA 478

RESULT 4
US-10-029-386-11381/c
Sequence 11381, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GE
TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
FILE REFERENCE: ABOMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 11381
LENGTH: 536
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL117379.14
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
OTHER INFORMATION: SWISSPROT HIT: P80398, EVALUATE 3.80e+00
OTHER INFORMATION: NT HIT: g114769835, EVALUATE 1.00e-78
OTHER INFORMATION: EST_HUMAN HIT: BF205331.1, EVALUATE 2.00e-78
US-10-029-386-11381

Query Match 4.1%; Score 117.8; DB 15; Length 536;
Best Local Similarity 83.2%; Pred. No. 4.4e-26;
Matches 134; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

OY 211 GCTTCAGTTTGGGACAGATATGATGATTAAGGACCTCTGAGCAATGAGAGACACC 270
DB 167 GCTTACACTTACCGGTAGGTATGAGAGCAAAAGGAGACCGAGCAATGAGAGGACT 108
OY 271 AAGGCTATCAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 330
DB 107 AAGGCTATCAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 48
OY 331 ATTGCCAAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 371
DB 47 ATGCCAAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7

RESULT 5

US-09-908-975-25239
Sequence 25239, Application US/09908975
Publication No. US20030165843A1
GENERAL INFORMATION:

APPLICANT: SHOSHAN, Avi
APPLICANT: WASSERMAN, Alon
APPLICANT: MINTZ, Eli
APPLICANT: MINTZ, Eli
TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
FILE REFERENCE: 36688-0005
CURRENT APPLICATION NUMBER: US/09/908,975
CURRENT FILING DATE: 2001-07-20
PRIOR APPLICATION NUMBER: US 60/287,724
PRIOR FILING DATE: 2001-05-02
PRIOR APPLICATION NUMBER: US 60/221,607
PRIOR FILING DATE: 2000-07-28
NUMBER OF SEQ ID NOS: 32337
SOFTWARE: PatentIn version 3.0
SEQ ID NO 25239
LENGTH: 65
TYPE: DNA
ORGANISM: Mus musculus
US-09-908-975-25239

Query Match 2.3%; Score 65; DB 10; Length 65;
Best Local Similarity 100.0%; Pred. No. 6.9e-10;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2331 TTGAAAGTCACTGTGCTGAGATCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2290
DB 1 TTGAAAGTCACTGTGCTGAGATCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 60

QY 2391 AGCAG 2395
Db 61 AGCAG 65

RESULT 6

US-09-764-864-471
Sequence 471, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT223
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 471
LENGTH: 915
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (894)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (903)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (905)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (914)
OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-471

Query Match 2.0%; Score 57.8; DB 9; Length 915;
Best Local Similarity 66.4%; Pred. No. 9.2e-07;

Matches 83; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1001 AGTGTAGATTACAGCCCAATGCCCTGTACTGATCGCCGACCTCACACACA 1060
Db 281 AGTGTAGATTACAGCCCAATGCCCTGTACTGATCGCCGACCTCACACACA 340
QY 1061 GATTATGATCTGCTGTGATGCTGTGAGAGTGTTCATGCTGTGTGGTATT 1120
Db 341 GATTATGATCTGCTGTGATGCTGTGAGAGTGTTCATGCTGTGTGGTATT 400
QY 1121 CTGAG 1125
Db 401 CTGAG 405

RESULT 7

US-10-175-523-142
Sequence 142, Application US/10175523
Publication No. US20030096264A1
GENERAL INFORMATION:
APPLICANT: Brockman, Jeffrey
APPLICANT: Evans, David
APPLICANT: Hook, Derek
APPLICANT: Klimczak, Leszek
APPLICANT: Laeng, Pascal
APPLICANT: Paley, Michael
APPLICANT: Rajan, Pithi
TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
FILE REFERENCE: 3235/13795-US3
CURRENT APPLICATION NUMBER: US/10/175,523
CURRENT FILING DATE: 2002-06-18
Prior application number: US 60/299,151
Prior filing date: 2001-06-18
Prior application number: US 60/317,828

PRIOR FILING DATE: 2001-09-07
PRIOR APPLICATION NUMBER: US 60/325,150
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/333,047
PRIOR FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: US 60/349,936
PRIOR FILING DATE: 2002-01-18
PRIOR APPLICATION NUMBER: US 60/361,834
PRIOR FILING DATE: 2002-03-04
NUMBER OF SEQ ID NOS: 197
SOFTWARE: PatentIn version 3.1
SEQ ID NO 142
LENGTH: 2320
TYPE: DNA
ORGANISM: Homo sapiens
US-10-175-523-142

Query Match 2.0%; Score 57.8; DB 15; Length 2320;
Best Local Similarity 66.4%; Pred. No. 1.8e-06;

Matches 83; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1001 AGTGTAGATTACAGCCCAATGCCCTGTACTGATCGCCGACCTCACACACA 1060
Db 109 AGTGTAGATTACAGCCCAATGCCCTGTACTGATCGCCGACCTCACACACA 168
QY 1061 GATTATGATCTGCTGTGATGCTGTGAGAGTGTTCATGCTGTGTGGTATT 1120
Db 169 GATTATGATCTGCTGTGATGCTGTGAGAGTGTTCATGCTGTGTGGTATT 228
QY 1121 CTGAG 1125
Db 229 CTGAG 233

RESULT 8

US-09-873-367C-112
Sequence 112, Application US/09873367C
Publication No. US20030165839A1
GENERAL INFORMATION:
APPLICANT: Young, Paul
APPLICANT: Soppet, Daniel
APPLICANT: Endress, Gregory
APPLICANT: Augustus, Meena
APPLICANT: Ebner, Reinhard
APPLICANT: Carter, Kenneth
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
FILE REFERENCE: 689290-64
CURRENT APPLICATION NUMBER: US/09/873,367C
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: U.S. 60/236,891
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: U.S. 60/236,842
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: U.S. 60/244,867
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: U.S. 60/245,084
PRIOR FILING DATE: 2000-11-01
NUMBER OF SEQ ID NOS: 1067
SOFTWARE: PatentIn version 3.0
SEQ ID NO 112
LENGTH: 5975
TYPE: DNA
ORGANISM: Homo sapiens
US-09-873-367C-112

Query Match 1.9%; Score 55.4; DB 10; Length 5975;
Best Local Similarity 59.0%; Pred. No. 2e-05;

Matches 95; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 1018 CCCAATGCCCTGTACTGATCGCCGACCTCACACAGATTATGATGCTCT 1077
Db 1200 CCCAATGCCCTGTACTGATCGCCGACCTCACACAGATTATGATGCTCT 1259

Query Match 1.9%; Score 55.4; DB 10; Length 6948;
Best Local Similarity 59.0%; Pred. No. 2.3e-05;
Matches 95; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 1018 CCCAATGCGCTGTAATGATGCTGCGCAGCTCAACAACAGGTTATGATGCTGT 1077
DB 2173 CCCAGCAAGAGTGTGGGTTTGGAAAAAACACATGCAACAGGTTATGCTGT 2232
QY 1078 GATGGGTGTGAGAGTGTTCATGTGTGATGTGTGGGTATTTCTGAGCCCGAGGGGG 1137
DB 2233 GGGAGATGTGATGATGCTGTTCATGTGTGATGTGTGGGTATTTCTGAGCAG 2292
QY 1138 CTCCTGGAAGAAACGGGGAAGACTACATCTGCCCAATG 1178
DB 2293 CAGATGGCGAGAGAGACAAAGATATGCTGTGTAATG 2333

RESULT 12
US-10-184-644-402
Sequence 402, Application US/10184644
Publication No. US20030044930A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C227
CURRENT APPLICATION NUMBER: US/10/184,644
CURRENT FILING DATE: 2002-06-28
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 402
LENGTH: 653
TYPE: PRT
ORGANISM: Homo Sapien
US-10-184-644-402

Query Match 1.8%; Score 50.4; DB 15; Length 653;
Best Local Similarity 6.6%; Pred. No. 0.00017;
Matches 43; Conservative 20; Mismatches 402; Indels 0; Gaps 0;

QY 618 GAAAGAACATCTGCTCTTCTGAAAAGCAAAAGAGGTGAAGAGAAAGACCTC 677
DB 2 DEDLLRQSLNCHGSSLSILRSEQDNPHFRSLGSAAPARPPQHPLOGKEXVDN 61
QY 678 TGACAGTGAAGTGAAGTGGCTTACGTTGAAGAACTTGAAACCGCTTGAGAAAGCG 737
DB 62 IEIQFTISKKADLLPALSWKSDPATSRINDESDHYAIMPLEQFMETPSMDRRLFFR 121
QY 738 AGAGCAAGAACCTGTGAGAGTGGCTTGAAGAGCACTGAGATGCTTGAGAAAGCG 797
DB 122 DIERGDIVIGRISIRFEPFWVLICLGGIMRDIHLBITLCPRLDVPSSHNDPLS 181
QY 798 CAGAGAGAAAGATTCTGCCGAAACTGGAGATGCCAAATAGAGAGTGGCAGAGACAG 857
DB 182 YVOTGDIITAGIKDIDRHEKLAIVLSYSSLPRLHSIGIKGLVYSBELPLYYRRVELNS 241
QY 858 ACCTCTCTTAAGAGAGCTGAGGCTGAGCTGTCAGGAGCAGAGTCCAGTCAGAGACAG 917
DB 242 NSLESTENVMOSLSLGVNPGVVFELKLGIDISNPPSLMRGLGKSPSEDDPASHLRK 301
QY 918 TCACATAGAAATTCAGTTGAAAGGAGGAGCTGAGGAAATATACAGAGGAAACCCGAG 977
DB 302 GQASVALKCVKIGVDYFVKGRHVDAKBNRYKALEIDKQNVBALVARGALVATGKINKAI 361

QY 978 GGAAGCGGCAACCAAGCCGAGTGTGAGTTACGACCCCAATGCTGTACTGAT 1037
DB 362 EDFELALENCPTRHNRKYLQTLVERGQLEBEKFLNABSYTKALALDSTFPDABEA 421
QY 1038 CTGCCGAGCCTCAACAACAGGTTATGATGCTGTGATGCTGTGAGAGTGT 1097
DB 422 LQKHKWQSLLEKQAEKBEKQTKLITSAKRLRLLEERLTKRKRKSTSSSV 481
QY 1098 CCATGTACTGTGTGGTATTTCTGAGCCCGAGGGCGCTCTCGAAAGAACGGGGA 1157
DB 482 SSADSESVSSSSSSGHRKHKRNRNSESRRSSRSNOIDQNRKDECTPVA 541
QY 1158 AGACTACATCTGCCCAATGACCAATTTTGCAAGTGAAGATGACAAACGTAAGCC 1217
DB 542 NTSASFNLHKGVEVRLKQDRLQYKQIKEMKCPSSSLEIPDPGVSYLPKUL 601
QY 1218 CACCAATGAGCAGACTGTGGTGAATCTGTGGTCTGATGCA 1264
DB 602 IKQPGSPGDIPEBGIVIIDDSIHVTDPEBQLQVGQDMEVDSGID 648

RESULT 13
US-10-184-634-402
Sequence 402, Application US/10184634
Publication No. US20030068684A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C217
CURRENT APPLICATION NUMBER: US/10/184,634
CURRENT FILING DATE: 2002-06-28
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 402
LENGTH: 653
TYPE: PRT
ORGANISM: Homo Sapien
US-10-184-634-402

Query Match 1.8%; Score 50.4; DB 15; Length 653;
Best Local Similarity 6.6%; Pred. No. 0.00017;
Matches 43; Conservative 20; Mismatches 402; Indels 0; Gaps 0;

QY 618 GAAAGAACATCTGCTCTTCTGAAAAGCAAAAGAGGTGAAGAGAAAGACCTC 677
DB 2 DEDLLRQSLNCHGSSLSILRSEQDNPHFRSLGSAAPARPPQHPLOGKEXVDN 61
QY 678 TGACAGTGAAGTGAAGTGGCTTACGTTGAAGAACTTGAAACCGCTTGAGAAAGCG 737
DB 62 IEIQFTISKKADLLPALSWKSDPATSRINDESDHYAIMPLEQFMETPSMDRRLFFR 121
QY 738 AGAGCAAGAACCTGTGAGAGTGGCTTGAAGAGCACTGAGATGCTTGAGAAAGCG 797
DB 122 DIERGDIVIGRISIRFEPFWVLICLGGIMRDIHLBITLCPRLDVPSSHNDPLS 181
QY 798 CAGAGAGAAAGATTCTGCCGAAACTGGAGATGCCAAATAGAGAGTGGCAGAGACAG 857
DB 182 YVOTGDIITAGIKDIDRHEKLAIVLSYSSLPRLHSIGIKGLVYSBELPLYYRRVELNS 241
QY 858 ACCTCTCTTAAGAGAGCTGAGGCTGAGCTGTCAGGAGCAGAGTCCAGTCAGAGACAG 917

Db 242 NSLSYENWQSSIGFVNPVVEFLLEKLGIDESPPLMGLQSKNPSBEDPAAALRK 301
Qy 918 TGACATAGAAAATCAGTTGAGAGGAGCGCATCGAGGAAATACAGAGAAAACCCCG 977
Db 302 QSASWALKCVKIGVDFVGRVDMNBNYKALBIDKQVVALVARGALYATKGSINRAI 361
Qy 978 GGAGCGGCGCAACCAAGCCCTGAGTGTGAGTTTACAGCCCAATGCCCTGTACTGAT 1037
Db 362 EDFLALNCPFHRRARXKLCOTLIVERGQGLEEBEKFMAEYKKALALDTEPDAEBA 421
Qy 1038 CTGCGCCGACCTCAGCAACAGAGTTTATGATCTGCTGTATCGGTGAGAGAGTGT 1097
Db 422 LQKHKYOKSLLEBKQAEKEEKOKTKITSAEKLRLLEBKLLKRRKRSSTSSSV 481
Qy 1098 CAGTGACTGTGAGTATTTCTAGAGCCCGGAGCGGCTCTCGTGAAGAGAGCGGGA 1157
Db 482 SSADSVSSSSSSSGHGRHKKRNRBSSSRSSRRSSSSNQIDNRDECPVPA 541
Qy 1158 AGACTACATCTGCCCAATTTGACCATTTTTCAGAGTGCAGATGAGACAAAGGTAAGCC 1217
Db 542 NTSASFVHKKQVEFKLGKQDRLQYEKTIQIKEDRCPLSSSLLEIPDDPGVYSYLFKXLT 601
Qy 1218 CACCATAGACGAGACTCTGGGTGAGATCTGTGGTGTCTGANTGCA 1264
Db 602 IKQPGAPSGDIPBEGVITIDDSIHVTPDPLQVGDMEVEDSGID 648

RESULT 14
US-10-085-783A-56748
; Sequence 56748, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: liev, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT FILING DATE: 2002-02-28
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/271,955
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 56748
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5)..(5)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (30)..(30)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-56748

Query Match 1.7%; Score 48; DB 13; Length 469;
Best Local Similarity 62.5%; Pred. No. 0.00076;
Matches 75; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 1059 CAGGTTTATGATCTGCTGTGATCGGTGAGAGTGTTCATGAGTGTGAGGTAT 1118
Db 116 CAGGTTTATGATCTGCTGTGATCGGTGAGAGTGTTCATGAGTGTGAGGTAT 1175
Qy 1119 TTCTGAGCGCGGAGCGGCTCTGGAAGAAAGGAGAAAGACTTACATCTGCCAAATTG 1178
Db 176 AAGTCTTCTCAAGCACAGCAGATGGCGGAGAAAGCAAGAAATATGTCTGTAAATG 235

RESULT 15
US-10-242-535A-56748
; Sequence 56748, Application US/10242535A
; Publication No. US2004001363A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: liev, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT FILING DATE: 2002-09-12
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 56748
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5)..(5)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (30)..(30)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-56748

Query Match 1.7%; Score 48; DB 16; Length 469;
Best Local Similarity 62.5%; Pred. No. 0.00076;
Matches 75; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
Search completed: April 23, 2004, 23:50:40
Job time: 1366.62 secs

Qy 1059 CAGGTTTATGATCTGCTGTGATCGGTGAGAGTGTTCATGAGTGTGAGGTAT 1118
Db 116 CAGGTTTATGATCTGCTGTGATCGGTGAGAGTGTTCATGAGTGTGAGGTAT 1175
Qy 1119 TTCTGAGCGCGGAGCGGCTCTGGAAGAAAGGAGAAAGACTTACATCTGCCAAATTG 1178
Db 176 AAGTCTTCTCAAGCACAGCAGATGGCGGAGAAAGCAAGAAATATGTCTGTAAATG 235

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 04:38:03 ; Search time 961.654 Seconds
(without alignment)
11529.921 Million cell updates/sec

Title: US-09-787-016a-1
Perfect score: 2610
Sequence: 1 ctgcgtgcgcgcgcgcac.....actcttaagatcatatcctg 2610

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapept 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001s:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2610	100.0	2610	AA290578	AA290578 Human dea
2	2516.2	96.4	2772	AAH15102	AAH15102 Human CDN
3	1810.4	69.4	2085	ABV28029	ABV28029 Human pro
4	1810.4	69.4	2085	ABV27935	ABV27935 Human pro
5	1810.4	69.4	2085	ABV22095	ABV22095 Human pro
6	1808.8	69.3	2085	ABV22192	ABV22192 Human pro
7	1564.4	59.9	1764	AACT77933	AACT77933 Human cen
8	1280.6	49.1	7290	AAK53352	AAK53352 Human pol
9	1280.6	49.1	7838	AAK52368	AAK52368 Human pol
10	1190.8	45.6	2332	AA886576	AA886576 DNA encod
11	1023.6	39.2	2867	AA290579	AA290579 Murine de
12	948.4	36.3	30626	AAK67051	AAK67051 Human imm
13	682.6	26.2	2986	AA886575	AA886575 DNA encod
14	596.8	22.9	659	AA808973	AA808973 Human bec
15	510.8	19.6	681	AAH08005	AAH08005 Human CDN
16	503	19.3	503	AA886574	AA886574 DNA encod
17	435	16.7	478	ACG31915	ACG31915 Human end
18	414.4	15.9	481	ABV42461	ABV42461 Human pro
19	336.6	12.9	542	AAH12970	AAH12970 Human CDN
20	314.2	12.0	362	ABV12528	ABV12528 Human pro
21	305.6	11.7	365	ABV42575	ABV42575 Human pro
22	305.6	11.7	365	ABV33664	ABV33664 Human pro
23	249.8	9.6	256	ABV03359	ABV03359 Human pro

24	235.2	9.0	408	AAA43780	AAA43780 Mouse sec
25	231.2	8.9	324	ABN21248	ABN21248 Human ORF
26	209	8.0	7838	AAK52368	AAK52368 Human pol
27	202.4	7.8	2310	AACT74383	AACT74383 Human sec
28	182.6	7.0	7174	ADG37602	ADG37602 Human nuc
29	137.2	5.3	936	ABV03329	ABV03329 Human pro
30	102.6	3.9	593	ABV12398	ABV12398 Human pro
31	101.2	3.9	179	ABV33540	ABV33540 Human pro
32	91.2	3.5	226	ABZ08454	ABZ08454 Human leu
33	78.2	3.0	6051	ABR29341	ABR29341 Drosophila
34	68.6	2.6	8751	ABL29340	ABL29340 Drosophila
35	65	2.5	5975	ABL61775	ABL61775 Colom ade
36	65	2.5	6256	AAIS8390	AAIS8390 Human pol
37	65	2.5	6256	ADB48359	ADB48359 Nov1 hum
38	65	2.5	6584	AAI60176	AAI60176 Human pol
39	65	2.5	6547	ACF34533	ACF34533 Gene enco
40	65	2.5	6948	AAH81806	AAH81806 Human dif
41	65	2.5	6996	AAK53025	AAK53025 Human pol
42	65	2.5	7354	AAK52041	AAK52041 Human pol
43	60	2.3	60	ABN33521	ABN33521 Human sp1
44	57.4	2.2	2000	ADA71938	ADA71938 Rice gene
45	56.8	2.2	10732	AAA10594	AAA10594 Gene enco

ALIGNMENTS

RESULT 1	AA290578	standard; cDNA; 2610 BP.
ID	AA290578	
XX	AA290578;	
AC		
XX		
DT	19-JUN-2000	(first entry)
XX		
DB	Human death inducer-oblierator 1 (DIO-1) polypeptide encoding cDNA.	
XX		
KW	Death inducer-oblierator 1; DIO-1; cell death; cancer; tumour; human; autoimmune disease; cytostatic; immunosuppressive; antidiabetic; antirheumatic; antiinflammatory; antiproliferative; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	265..1953
FT		/*tag= a
FT		/product= "DIO-1"
XX		
PN	WO200015787-A1.	
XX		
PD	23-MAR-2000.	
XX		
PF	10-SEP-1999;	99WO-GB003019.
XX		
PR	10-SEP-1998;	98SE-00003069.
XX		
PR	17-SEP-1998;	98US-0100873P.
XX		
PA	(CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.	
PA	(BANN/) BANNERMAN D G.	
PI	Alonso CM, Domingo DG, Grandien A, Leonardo B, Martinez P;	
XX		
DR	WPI: 2000-271426/23.	
DR	P-PSDB: AAY67579.	
XX		
PT	New DNA encoding human and murine death inducer-oblierator 1	
PT	polypeptides, useful in the treatment of cancer, autoimmune diseases,	
PT	diabetes, rheumatoid arthritis, benign tumors, malignant tumors and	
PT	hyperproliferative skin disorders.	
XX		
PS	Claim 1; Fig 1A; 27pp; English.	
XX		
CC	The invention provides nucleic acids encoding the human and murine death	


```
Db 1921 TGTCCAGGACTGGAGTTGACGTTTGTGTAACTGATCAGACACCGGCTGCACAT 1980
XX
Qy 1981 CAGCGGGAAGACGAGCCCATGTCAGAGATGCTCCCTGCTGCTGTCATCCCAAGTC 2040
Db 1981 CAGCGGGAAGACGAGCCCATGTCAGAGATGCTCCCTGCTGCTGTCATCCCAAGTC 2040
Qy 2041 TGTGAGAGCTTCTGTCTGACTGTTTTCGAAAGCTGTAAAGCTGAGTGAAGTTCACCT 2100
Db 2041 TGTGAGAGCTTCTGTCTGACTGTTTTCGAAAGCTGTAAAGCTGAGTGAAGTTCACCT 2100
Qy 2101 TAAATGATGATTTCTTAAATCTCTGTTTTCAGTCTGAGGCTGTGTAAGTATTGATTC 2160
Db 2101 TAAATGATGATTTCTTAAATCTCTGTTTTCAGTCTGAGGCTGTGTAAGTATTGATTC 2160
Qy 2161 CTTCATCTCCAGTCTGATGATGACCACTGCCCCGCGACGCACTCCACCCCTGCTG 2220
Db 2161 CTTCATCTCCAGTCTGATGATGACCACTGCCCCGCGACGCACTCCACCCCTGCTG 2220
Qy 2221 CACATGAGTGTCTGACAAACAGCGCTGTATAGCTTCAGTTCCTTCACATGTCACCG 2280
Db 2221 CACATGAGTGTCTGACAAACAGCGCTGTATAGCTTCAGTTCCTTCACATGTCACCG 2280
Qy 2281 CCAGCAGATGAAGAATCATCTTTTGTATGTTGTGGAATCTTTGCAAGTTAGTGTG 2340
Db 2281 CCAGCAGATGAAGAATCATCTTTTGTATGTTGTGGAATCTTTGCAAGTTAGTGTG 2340
Qy 2341 CATCTGATTTTCAGGTGTATCATTTATTTTGACTGGGAGATGAGGGATTTTTTTTTT 2400
Db 2341 CATCTGATTTTCAGGTGTATCATTTATTTTGACTGGGAGATGAGGGATTTTTTTTTT 2400
Qy 2401 TCCATGTCGATTCACACGCTACACACCCACATGAAACATTCGAAAGGACAC 2460
Db 2401 TCCATGTCGATTCACACGCTACACACCCACATGAAACATTCGAAAGGACAC 2460
Qy 2461 ACTCTGCTTCATGAGGCCCCACGCTAAGTGAAGTTCACACTGAGAACACTGTCTGACCC 2520
Db 2461 ACTCTGCTTCATGAGGCCCCACGCTAAGTGAAGTTCACACTGAGAACACTGTCTGACCC 2520
Qy 2521 AGGAGCGCTGCTTGAAGCTTGTATTTTCAATGTAATGCTGCTGCTGCTGCTGCT 2580
Db 2521 AGGAGCGCTGCTTGAAGCTTGTATTTTCAATGTAATGCTGCTGCTGCTGCTGCT 2580
Qy 2581 GAATGTTTGAATCTTGAATGATATCTG 2610
Db 2581 GAATGTTTGAATCTTGAATGATATCTG 2610

RESULT 2
AAH15102
ID AAH15102 standard; cDNA; 2772 BP.
XX
AC AAH15102;
XX
AC 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:13129.
XX
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN BP1074617-A2.
XX
PD 07-FEB-2001.
XX
PP 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
```

```
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 13129; 2537bp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification, where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 2772 BP; 699 A; 691 C; 765 G; 617 T; 0 U; 0 Other;
XX
Query Match 96.4%; Score 2516.2; DB 4; Length 2772;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2541; Conservative 0; Mismatches 3; Indels 3; Gaps 2;
XX
Qy 65 GAGGCTTACTCCAGCGGGAACAGCTCTAGATATCTGAGTTGTGAAATACGAAGCTG 124
Db 101 GAGGCTTACTCCAGCGGGAACAGCTCTAGATATCTGAGTTGTGAAATACGAAGCTG 160
Qy 125 TTAATCTGTAACAGTGGCTGACAAAGTGTGTGAGCTGCTGCTGCTTGAACC 184
Db 161 TTAATCTGTAACAGTGGCTGACAAAGTGTGTGAGCTGCTGCTGCTTGAACC 220
Qy 185 AAGAGTTTGTCTGCGACAGGTTTGTGTGATTTAGATTCAGGGAAGAGTGTCAAG 244
Db 221 AAGAGTTTGTCTGCGACAGGTTTGTGTGATTTAGATTCAGGGAAGAGTGTCAAG 280
Qy 245 CTTCAGTGTGAGAGGATATGACGAAAGGCGACCCGACCAATGAGAGGCACTTA 304
Db 281 CTTCAGTGTGAGAGGATATGACGAAAGGCGACCCGACCAATGAGAGGCACTTA 340
Qy 305 AGGCAATCAACCCACAGCAAAAGATTGAGAAAACATGGGGTTTTCAGAGCACTTA 364
Db 341 AGGCAATCAACCCACAGCAAAAGATTGAGAAAACATGGGGTTTTCAGAGCACTTA 400
Qy 365 TGGCCAAAGGAGAGGGGCGAGGGGAGCGGAGGCTGACCACTGAGAGCGCCACCCAC 424
Db 401 TGGCCAAAGGAGAGGGGCGAGGGGAGCGGAGGCTGACCACTGAGAGCGCCACCCAC 460
Qy 425 AGCAGAGCTGGGCTGTCTGCTGCGCGCAGTGGAGAGGACCCAGCGCACTGAGCGG 484
Db 461 AGCAGAGCTGGGCTGTCTGCTGCGCGCAGTGGAGAGGACCCAGCGCACTGAGCGG 520
Qy 485 TGGAGAGTTCTGATCACTTGGCGGGCGCGCGGCGAGAGAGCAATGCTGTCTCTCTGG 544
```

D 521 TGGAGCAGTTCTTGACCAATTGGGGGCGCGGCGAGGAGGACATGCTGTCTCCCTGG 580
Q 545 AGGATTCGTGTAGCCCACTGCTGCCCCGCA CAGACGCCGAGACAGCTCTCCAGGCGCA 604
D 581 AGGATTCGTGTAGCCCACTGCTGCCCCGCA CAGACGCCGAGACAGCTCTCCAGGCGCA 640
Q 605 GCGGTGAAGGCGCTTCTGAGACCAAGAGCGGCCCCAGCTGTCTTCCAGCTGTAGAG 664
D 641 GCGGTGAAGGCGCTTCTGAGACCAAGAGCGGCCCCAGCTGTCTTCCAGCTGTAGAG 700
Q 665 AACGACCAAGCTCTTCTGAAAAAGTGAAGAGGAGATGACACGATGACACTTCCGATA 724
D 701 AACGACCAAGCTCTTCTGAAAAAGTGAAGAGGAGATGACACGATGACACTTCCGATA 760
Q 725 GTGACAGGATGCGCTGACTTGTAAAGAGCTTCAAGATCCCTTGTGCGAGAGCGCGAAC 784
D 761 GTGACAGGATGCGCTGACTTGTAAAGAGCTTCAAGATCCCTTGTGCGAGAGCGCGAAC 820
Q 785 AGGAGCCCACTGAGAGGCGCCCTGAAAAAGGATCCAGAGTCGCTGCGAGAGAGCGCGAG 844
D 821 AGGAGCCCACTGAGAGGCGCCCTGAAAAAGGATCCAGAGTCGCTGCGAGAGAGCGCGAG 880
Q 845 AGGAGGTCCTCCGCGAGACTGTGCGCTCCAGGCCAGTGAACA CTGTGAGAGGCGCTCTGC 904
D 881 AGGAGGTCCTCCGCGAGACTGTGCGCTCCAGGCCAGTGAACA CTGTGAGAGGCGCTCTGC 940
Q 905 CCAAGTAAAGAGAGCCCGAGAACGATCAAGGGGTTGTGTCCAGGCTGGAAAAATGACA 964
D 941 CCAAGTAAAGAGAGCCCGAGAACGATCAAGGGGTTGTGTCCAGGCTGGAAAAATGACA 1000
Q 965 GAGAGAGTAAGTTGAGAGGAAAGCGGCTCAGACATCAAGATGAGAGGCTGTGAGACT 1024
D 1001 GAGAGAGTAAGTTGAGAGGAAAGCGGCTCAGACATCAAGATGAGAGGCTGTGAGACT 1060
Q 1025 TGGGCGGACCGAAGCTGTAATGTGAGGGTTACGACCCCAAGCGCTGTATTGCAATTTGCG 1084
D 1061 TGGGCGGACCGAAGCTGTAATGTGAGGGTTACGACCCCAAGCGCTGTATTGCAATTTGCG 1120
Q 1085 GCCAGCTCACAACAACAGGTTTATGATTTGTGTGACCGCTGTGAAAGATGTTTCTATG 1144
D 1121 GCCAGCTCACAACAACAGGTTTATGATTTGTGTGACCGCTGTGAAAGATGTTTCTATG 1180
Q 1145 GCGATGTGTGGGCACTTTCTGAGGCTCGAGGGAGGCTTTTGAAGGAATGAGGAAAGACT 1204
D 1181 GCGATGTGTGGGCACTTTCTGAGGCTCGAGGGAGGCTTTTGAAGGAATGAGGAAAGACT 1240
Q 1205 ATATCTGCCCAAACTGCAACAATTCTGCAAGTGCAGAGTGAAGACTATTCCAGAAACGGCAG 1264
D 1241 ATATCTGCCCAAACTGCAACAATTCTGCAAGTGCAGAGTGAAGACTATTCCAGAAACGGCAG 1300
Q 1265 ATCAGCAGGAAGCTTAATGAGAGCTTGAAGATGTGATGAGCAGCAATTGTATACAAATTAG 1324
D 1301 ATCAGCAGGAAGCTTAATGAGAGCTTGAAGATGTGATGAGCAGCAATTGTATACAAATTAG 1360
Q 1325 GAAACAATGAGAGAGGCTGAGCGAAGCAAGGGATTAAGGGTTGAATTGAGAAAGCTG 1384
D 1361 GAAACAATGAGAGAGGCTGAGCGAAGCAAGGGATTAAGGGTTGAATTGAGAAAGCTG 1420
Q 1385 CAAATCCAAAGTGCAGAAAGAACTCAAGATCTTCCAGCTGTGATGAGAGCGCTGTGTG 1444
D 1421 CAAATCCAAAGTGCAGAAAGAACTCAAGATCTTCCAGCTGTGATGAGAGCGCTGTGTG 1480
Q 1445 CCTCAAAATGTATTTGGGCGCGGCTGTCTCACTGTGCGCAGCCGACCTGAGTATCTGCA 1504
D 1481 CCTCAAAATGTATTTGGGCGCGGCTGTCTCACTGTGCGCAGCCGACCTGAGTATCTGCA 1540
Q 1505 GTAATGATCTGTATCTCTCAAAACGCGCGAGCAGCAATGAAAGTTTCTAAGCTCAGGTTAAG 1564
D 1541 GTAATGATCTGTATCTCTCAAAACGCGCGAGCAGCAATGAAAGTTTCTAAGCTCAGGTTAAG 1600
Q 1565 AACAGAGCCCAAAGCTTAAGAAAAAGATGAAGTGAAGCAGAGAGCCCAAGCTCTTCCGA 1624
D 1601 AACAGAGCCCAAAGCTTAAGAAAAAGATGAAGTGAAGCAGAGAGCCCAAGCTCTTCCGA 1660

Q 1625 AATGGGGTGTCAAGCAGGTAATTAATCTCTTCTGTGACCAAGAGACAGCTCCAGAAA 1684
D 1661 AATGGGGTGTCAAGCAGGTAATTAATCTCTTCTGTGACCAAGAGACAGCTCCAGAAA 1720
Q 1685 AAAAAAGACCAAGTGAAGAGCAAGTGTGTCTCCGCGGAGTGAAGCACTCGGGA 1744
D 1721 AAAAAAGACCAAGTGAAGAGCAAGTGTGTCTCCGCGGAGTGAAGCACTCGGGA 1780
Q 1745 AGGAAGCAGCTTGTGAGAGAGCAAGCGGTGTGTGGGCGAGATGACAAATTACATGAG 1804
D 1781 AGGAAGCAGCTTGTGAGAGAGCAAGCGGTGTGTGGGCGAGATGACAAATTACATGAG 1840
Q 1805 TAAAGCCAGAAAAAGCTGCTGCTCCCTGCGGCTCACTGTGTATTAATGATGATCAAC 1864
D 1841 TAAAGCCAGAAAAAGCTGCTGCTCCCTGCGGCTCACTGTGTATTAATGATGATCAAC 1900
Q 1865 TAGGGGTTGGCTCTGTGACCCCTCCGCTTCTTGTGATAGCCATCCCTGAGGCTGTG 1924
D 1901 TAGGGGTTGGCTCTGTGACCCCTCCGCTTCTTGTGATAGCCATCCCTGAGGCTGTG 1960
Q 1925 CAGACCTGGAGTTGCAAGCTTGTGTGTTAAGCTGAATCAAGACACCGGCGACATCAGC 1984
D 1961 CAGACCTGGAGTTGCAAGCTTGTGTGTTAAGCTGAATCAAGACACCGGCGACATCAGC 2020
Q 1985 GGAAGACAGAGCCCAATGTCCAGAGTGCCTCTGCTGCGCTGTGTCCATCCCTAGTCTGT 2044
D 2021 GGAAGACAGAGCCCAATGTCCAGAGTGCCTCTGCTGCGCTGTGTCCATCCCTAGTCTGT 2080
Q 2045 AGGACTTCTGTCACTGTTTTCAAAGCTGTAACTGACCTGATGAAGTTGACTTAT 2104
D 2081 AGGACTTCTGTCACTGTTTTCAAAGCTGTAACTGACCTGATGAAGTTGACTTAT 2140
Q 2105 GATTGATCTTAAATCTCTGTTTTCACCTGAGGCTCTGTAAGATTTGATTTCTCTC 2164
D 2141 GATTGATCTTAAATCTCTGTTTTCACCTGAGGCTCTGTAAGATTTGATTTCTCTC 2200
Q 2165 ATCCAGCTGATTTGATGACACACTGCGCGGACGCGACATCCACCTGTGTGACCA 2224
D 2201 ATCCAGCTGATTTGATGACACACTGCGCGGACGCGACATCCACCTGTGTGACCA 2260
Q 2225 TGAATGTTCTGACCAACGCGCTGTAAAGCTTCAAGTTTTCACATGTGCCAGGCCAG 2284
D 2261 TGAATGTTCTGACCAACGCGCTGTAAAGCTTCAAGTTTTCACATGTGCCAGGCCAG 2320
Q 2285 CACATGAAGACATCACTCTTTTATGTTGAGGAATCTTGAAGTTAGTTGACATC 2344
D 2321 CACATGAAGACATCACTCTTTTATGTTGAGGAATCTTGAAGTTAGTTGACATC 2380
Q 2345 TGAATTTGAGGTTGACATTTATTTTGACTGGGACAGATGAGGAAATTTTTTTTCCCA 2404
D 2381 TGAATTTGAGGTTGACATTTATTTTGACTGGGACAGATGAGGAAATTTTTTTTCCCA 2438
Q 2405 TGTCCGATTCACACCTTACACACCACTGAACACATTTGAACTTGAAG--CAACACT 2463
D 2439 TGTCCGATTCACACCTTACACACCACTGAACACATTTGAACTTGAAG--CAACACT 2498
Q 2464 CCTGCTTCAATAGAGCCCAACGTAAGTGTCAACCTAAGAACATGATCCGACGAGG 2523
D 2499 CCTGCTTCAATAGAGCCCAACGTAAGTGTCAACCTAAGAACATGATCCGACGAGG 2558
Q 2524 ACGGCTGCTTGAACCTTGTGATTTCAACATGTAAGTGTGCTTCTGCTCTTGA 2583
D 2559 ACGGCTGCTTGAACCTTGTGATTTCAACATGTAAGTGTGCTTCTGCTCTTGA 2618
Q 2584 TGTTTAGAATCTTTAAGATCATATCTG 2610
D 2619 TGTTTAGAATCTTTAAGATCATATCTG 2645

RESULT 3
ABV28029
ID ABV28029 standard; cDNA; 2085 BP.

XX AC ABV28029;
XX DT 16-SBP-2002 (first entry)
XX DB Human prostate expression marker cDNA 28020.
XX KM Human; prostate cancer; cytosolic; carcinogen; pharmacodynamic marker;
XX KM pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
XX PN W0200160860-A2.
XX PD 23-AUG-2001.
XX PF 20-FEB-2001; 2001WO-US005171.
XX PR 17-FEB-2000; 2000US-0183319P.
XX PR 16-MAR-2000; 2000US-0189862P.
XX PR 25-MAY-2000; 2000US-0207454P.
XX PR 09-JUN-2000; 2000US-0211314P.
XX PR 18-JUL-2000; 2000US-0219007P.
XX PR 13-DEC-2000; 2000US-0255281P.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Schlegel R, Endege WO, Monahan JB;
XX DR WPI; 2001-662795/76.
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer, useful
XX PT for detecting presence of prostate cancer, stage of prostate cancer.
XX Claim 1; Page 5788; 11750BP; English.
XX The invention relates to an isolated nucleic acid molecule (I) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX specification or its complement. (I) is useful for: (a) assessing whether
XX a patient is afflicted with prostate cancer; (b) monitoring the efficacy
XX of a test compound to inhibit prostate cancer in a patient; (c) assessing the
XX progression of prostate cancer in a patient; (d) assessing the efficacy
XX of a test compound to inhibit prostate cancer in a patient; (e) assessing
XX the efficacy of a composition for inhibiting prostate cancer in a patient;
XX (f) assessing the prostate cell carcinogenic potential of a compound; (g)
XX determining whether prostate cancer has metastasized in a patient; (h)
XX assessing the aggressiveness or indolence of prostate cancer in a patient
XX ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
XX Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;
XX
XX Query Match 69.4%; Score 1810.4; DB 5; Length 2085;
XX Best Local Similarity 98.8%; Pred. No. 0;
XX Matches 1960; Conservative 0; Mismatches 11; Indels 13; Gaps 13;
XX
XX 3 CGGTGCGCGTCCGCGCCACTCCGCGCGTTCGCGGAATGCTGCGAGACCTAGAGGCT 62
XX |||||
XX 17 CGGTGCGCGTCCGCGCCACTCCGCGCGTTCGCGGAATGCTGCGAGACCTAGAGGCT 76
XX |||||
XX 63 GCGGAGCTTACTCCAGCGGAAACAGCTCTAGATTAATCTGAGTTGTGAAATACAGAGCC 122
XX |||||
XX 77 GCGGAGCTTACTCCAGCGGAAACAGCTCTAGATTAATCTGAGTTGTGAAATACAGAGCC 135
XX |||||
XX 123 TGTATCTGTGAACAGTGTGCTGCAACAGTGTGTGAGAGCTGCTGTGCTTGAGC 182
XX |||||
XX 136 TGTATCTGTGAACAGTGTGCTGCAACAGTGTGTGAGAGCTGCTGTGCTTGAGC 195
XX |||||
XX 183 CCAGAGTTTGTCTGTGCAAGGTTTGTGTTATTTAGATTTCAGGGAAGTGTCCA 242
XX |||||
XX 196 CCAGAGTTTGTCTGTGCAAGG- TTTGTGTTTATTTAGATTTCAGGGAAGTGTCCA 254
XX |||||
XX 243 AGCTTTCAGTGTGAGAGAGTATGAGACGAAAGGCGACCCGAGCATGAGAGGACC 302
XX |||||

DB 255 AGCTTTCAGTGTGAGAGAGTATGAGACGAAAGGCGACCCGAGCATGAGAGGACC 313
XX |||||
XX 303 TAAGGCGCATCAAAACCCAGCAGCAAAAGTTTCAGGAAAACATGGGGTTTTCGAGAGCAC 362
XX |||||
XX 314 TAAGGCGCATCAAAACCCAGCAGCAAAAGTTTCAGGAAAACATGGGGTTTTCGAGAGCAC 373
XX |||||
XX 363 TATGCGCAAGGAGAGGGGCGCAAGGAGTCCGAGGCTGACCTGAGAGCCGACCC 422
XX |||||
XX 374 TATGCGCAAGGAGAGGGGCGCAAGGAGTCCGAGGCTGACCTGAGAGCCGACCC 432
XX |||||
XX 423 ACAAGAGAGCGTGGGCGCTGCTCCGCGGCGAGTGGGAGGCGCCAGAGCGCATGAGG 482
XX |||||
XX 433 ACAAGAGAGCGTGGGCGCTGCTCCGCGGCGAGTGGGAGGCGCCAGAGCGCATGAGG 492
XX |||||
XX 483 CGTGAAGAGTTCTGACCAATTCGCGGCGCGCGCGAGAGAGCAATGCTGTCTCCT 542
XX |||||
XX 493 CGTGAAGAGTTCTGACCAATTCGCGGCGCGCGCGAGAGAGCAATGCTGTCTCCT 552
XX |||||
XX 543 GAGAGATTCTGCTGAGCCGACGCTCTGCGCGCGAGAGAGCGCGAGAGAGAGCTCGAGG 602
XX |||||
XX 553 GAGAGATTCTGCTGAGCCGACGCTCTGCGCGCGAGAGAGCGCGAGAGAGAGCTCGAGG 610
XX |||||
XX 603 CAGCGTGAAGAGCGCTTCTGAGACCGAGAGGCGCGCGAGAGAGAGCTTCTGAGAGTGA 662
XX |||||
XX 611 CAGCGTGAAGAGCGCTTCTGAGACCGAGAGGCG-CCGAGTGTCTTCCAGAG-TGTGA 668
XX |||||
XX 663 GGAACGACCAAGCGCTCTTCTGAGAAAGGTGAAGAGAGGAGTGAACAGATGACCTCCGA 722
XX |||||
XX 669 GGAACGACCAAG-CTCTTCTGAGAAAGGTGAAGAGAGGAGTGAACAGATGACCTCCGA 727
XX |||||
XX 723 TAGTGAAGAGCGAGTGCCTGACCTTGAAGAGCTTGAAGATGCTTCTGAGAGAGCGGA 782
XX |||||
XX 728 TAGTGAAGAGCGAGTGC-CTGACCTTGAAGAGCTTGAAGATGCTTCTGAGAGAGCGGA 785
XX |||||
XX 783 ACAAGAGCGCATGAGAGGCGCTTGAAGAGGATTCAGAGTGCCTGCGGAGAGAGCGCG 842
XX |||||
XX 786 ACAAGAGCGCATGAGAGGCGCTTGAAGAGGATTCAGAGTGCCTGCGGAGAGAGCGCG 845
XX |||||
XX 843 GAGAGAGGCTCCGCGCGAGACTGTGGGCTCCGAGGCGCAGTGAACATGTCGAGAGGCTCT 902
XX |||||
XX 846 GAGAGAGGCTCCGCGCGAGACTGTGGGCTCCGAGGCGCAGTGAACATGTCGAGAGGCTCT 905
XX |||||
XX 903 GCCAGTAAGCAGAGAGCCGAGAGCAGTCAAGGAGGCTGTGTCTCCAGGCTGGAAGATGA 962
XX |||||
XX 906 GCCAGTAAGCAGAGAGCCGAGAGCAGTCAAGGAGGCTGTGTCTCCAGGCTGGAAGATGA 965
XX |||||
XX 963 CAGAGAGGTAAGTTGAGAGGAAAGGCGCTCAGAGCATCAAGATGAGAGGCTGAGAG 1022
XX |||||
XX 966 CAGAGAGGTAAGTTGAGAGGAAAGGCGCTCAGAGCATCAAGATGAGAGGCTGAGAG 1025
XX |||||
XX 1023 CTTGGGCGGACCGAAGCCTGAATGTAGGGTTACAGACCCCAAGCGCCTGTATTCATTGG 1082
XX |||||
XX 1026 CTTGGGCGGACCGAAGCCTGAATGTAGGGTTACAGACCCCAAGCGCCTGTATTCATTGG 1085
XX |||||
XX 1083 CCGCAGCCTCAACAACAAGGTTATGATTTGCTGTGAACCCCTGAGAGAAATGCTTTCA 1142
XX |||||
XX 1086 CCGCAGCCTCAACAACAAGGTTATGATTTGCTGTGAACCCCTGAGAGAAATGCTTTCA 1145
XX |||||
XX 1143 TGGCGATTTGTGGGCAATTCGAGGCTCGAGGAGGCTTTTGAAGAAATGCGGAGAG 1202
XX |||||
XX 1146 TGGCGATTTGTGGGCAATTCGAGGCTCGAGGAGGCTTTTGAAGAAATGCGGAGAG 1205
XX |||||
XX 1203 CTATATCTGCCAAACTGACCAATTCGAGAGTGAAGATGAGACTCATTCAGAAAACGCG 1262
XX |||||
XX 1206 CTATATCTGCCAAACTGACCAATTCGAGAGTGAAGATGAGACTCATTCAGAAAACGCG 1265
XX |||||
XX 1263 AGATGAGCAGAGAGCTAAATGAGAGCTGAGAGTGTGATGCGACCGATTTGACAGTAT 1322
XX |||||
XX 1266 AGATGAGCAGAGAGCTAAATGAGAGCTGAGAGTGTGATGCGACCGATTTGACAGTAT 1325
XX |||||
XX 1323 AGGAACATGAGAGCAGAGTCTAGGAGAGACCAAGGATTAAGGGTGAATGAGAAACG 1382
XX |||||

Oy 603 CAGCGTGAAGACGCTTCTGAGACCAAGACGCGCCCGCAGTCTGCTTCCACGCTGTAA 662
 |||||
 Db 611 CAGCGTGAAGACGCTTCTGAGACCAAGACGCG -CCCCAGTCTGCTTCCACG -TGTAA 668
 |||||
 Oy 663 GGAACGACGACCTTTCTGAAAAAGTGAAGAGGGGATGACCAAGATGACCACTCCGA 722
 |||||
 Db 669 GGAACGACGAC -CTTCTTCTGAAAAAGTGAAGAGGGGATGACCAAGATGACCACTCCGA 727
 |||||
 Oy 723 TAGTGAACAGGATGAGCTGACCTTGAAGAGCTTCAAGATCGCTTCCAGAGAACGGGA 782
 |||||
 Db 728 TAGTGAACAGGATGAG -CTGACCTTGAAGAGCTTCAAGATCG -CTTCCAGAGAACGGGA 785
 |||||
 Oy 783 ACAGAGACCCCACTGAGAGAGCCCTGTAAGAGATCCAGAGTGCCTGCGAGAGAGCGCG 842
 |||||
 Db 786 ACAGAGACCCCACTGAGAGAGCCCTGTAAGAGATCCAGAGTGCCTGCGAGAGAGCGCG 845
 |||||
 Oy 843 GAGAGAGGCTCCCGCGGAGACTGTGGGCTCCGAGGCGAGTGCACCTGTGAGAGGGGCTCT 902
 |||||
 Db 846 GAGAGAGGCTCCCGCGGAGACTGTGGGCTCCGAGGCGAGTGCACCTGTGAGAGGGGCTCT 905
 |||||
 Oy 903 GCCCAGTAAGAGAGAGCCCGAGAGACATCAAGGGGTTGTCTCCAGGCTGGAGAAAGATGA 962
 |||||
 Db 906 GCCCAGTAAGAGAGAGCCCGAGAGACATCAAGGGGTTGTCTCCAGGCTGGAGAAAGATGA 965
 |||||
 Oy 963 CAGAGAGATGAAGTTGAG 1022
 |||||
 Db 966 CAGAGAGATGAAGTTGAG 1025
 |||||
 Oy 1023 CTTGGGCGGAGCGAGAGAGCTGAAATGTGAGAGGTTACGACCCCAAGCGCTGTATTTGCA 1082
 |||||
 Db 1026 CTTGGGCGGAGCGAGAGCTGAAATGTGAGAGGTTACGACCCCAAGCGCTGTATTTGCA 1085
 |||||
 Oy 1083 CGGCGAGGCTTCAACAACAAGGTTATGATTTGCTGTGACCGCTGTGAGAGAGGTTTCA 1142
 |||||
 Db 1086 CGGCGAGGCTTCAACAACAAGGTTATGATTTGCTGTGACCGCTGTGAGAGAGGTTTCA 1145
 |||||
 Oy 1143 TGGCGATTGTGGGCAATTTCTGAGGCTTGAAGAGGCTTTTGAAGAGAGATGGGAGAGA 1202
 |||||
 Db 1146 TGGCGATTGTGGGCAATTTCTGAGGCTTGAAGAGGCTTTTGAAGAGAGATGGGAGAGA 1205
 |||||
 Oy 1203 CTATATCTGCCCCAATCTGACCAATCTGCAAGTGAAGAGATGAGACATTCAGAAACGCG 1262
 |||||
 Db 1206 CTATATCTGCCCCAATCTGACCAATCTGCAAGTGAAGAGATGAGACATTCAGAAACGCG 1265
 |||||
 Oy 1263 AGATCAGCAGAGAGCTTAATGAGAGACTGAGATGCTGATGAGACCGATTTGTACAAGTAT 1322
 |||||
 Db 1266 AGATCAGCAGAGAGCTTAATGAGAGACTGAGATGCTGATGAGACCGATTTGTACAAGTAT 1325
 |||||
 Oy 1323 AGGAAACAATAGAGCAGAAAGTCTAGCGAAGACCAAGAGATTAAGGGTGAATTTGAGAAAGC 1382
 |||||
 Db 1326 AGGAAACAATAGAGCAGAAAGTCTAGCGAAGACCAAGAGATTAAGGGTGAATTTGAGAAAGC 1385
 |||||
 Oy 1383 TGCAGATTCAGAGTGGCAAGAGAACTCAAGATCTTCCAGGCTGTGATGAGAGCGGCTTGG 1442
 |||||
 Db 1386 TGCAGATTCAGAGTGGCAAGAGAACTCAAGATCTTCCAGGCTGTGATGAGAGCGGCTTGG 1445
 |||||
 Oy 1443 TGCCTCAAAATATATATGAGCCCGGGTGTGCTGACAGTGGGCGAGCCCGAATCGGTGATCTG 1502
 |||||
 Db 1446 TGCCTCAAAATATATATGAGCCCGGGTGTGCTGACAGTGGGCGAGCCCGAATCGGTGATCTG 1505
 |||||
 Oy 1503 CAGTAATGACTGTATCTTCAAAACAGCGCGCAGCGACAATGAAGTTTCTAAGCTCAGTTAA 1562
 |||||
 Db 1506 CAGTAATGACTGTATCTTCAAAACAGCGCGCAGCGACAATGAAGTTTCTAAGCTCAGTTAA 1565
 |||||
 Oy 1563 AGAACAAGAGCCAAAGCTTAAGAGAAATGAGAGTGAAGCCCAAGAGAGCCCAAGCTTCC 1622
 |||||
 Db 1566 AGAACAAGAGCCAAAGCTTAAGAGAAATGAGAGTGAAGCCCAAGAGAGCCCAAGCTTCC 1625
 |||||
 Oy 1623 GAAATGCGGTGCTCAGGAGAGTATTAATCTCTTCTGTGCAACAAGAGACGCTCCAGA 1682
 |||||
 Db 1626 GAAATGCGGTGCTCAGGAGAGTATTAATCTCTTCTGTGCAACAAGAGACGCTCCAGA 1685
 |||||
 Oy 1683 AAAAAAAGAGACCAAGTGAAGAGAGCAGTGTGTCTCCTGCGGAGTGAAGACACTCGG 1742
 |||||

Db 1686 AAAAAAAGAGACCAAGTGAAGAGAGCAGTGTGTCTCCTGCGGAGTGAAGACACTCGG 1745
 |||||
 Oy 1743 GAGAGAGCAGCTTGTGAG 1802
 |||||
 Db 1746 GAGAGAGCAGCTTGTGAG 1805
 |||||
 Oy 1803 AGTAAGCCAGAGAAAGACTGTGCTCTCCCTGCTGACAGTGTGTATTAATGTATATCA 1862
 |||||
 Db 1806 AGTAAGCCAGAGAAAGACTGTGCTCTCCCTGCTGACAGTGTGTATTAATGTATATCA 1865
 |||||
 Oy 1863 CCTAGGGGTGGGCTCTCGAGACCCCTCCCGTCTTTTGTGATAGACATCCCTGGGCTG 1922
 |||||
 Db 1866 CCTAGGGGTGGGCTCTCGAGACCCCTCCCGTCTTTTGTGATAG -CATCCCTGGGCTG 1924
 |||||
 Oy 1923 TCCAGAGACTGGAGATTGACGCTTGTGTAACTGATCAACAACCGGCTGACACATCA 1982
 |||||
 Db 1925 TCCAGAGACTGGAGATTGACG -TTGTGTAACTGATCAACAACCGGCTGCGGGA 1983
 |||||
 Oy 1983 GCGG 1986
 |||||
 Db 1984 GCAG 1987
 |||||
 RESULT 5
 ABV22095
 ID ABV22095 standard; cDNA; 2085 BP.
 XX
 AC ABV22095;
 DT 13-SEP-2002 (first entry)
 XX
 DB Human prostate expression marker cDNA 22086.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US005171.
 XX
 PR 17-FEB-2000; 2000US-0183319P.
 PR 16-MAR-2000; 2000US-0189862P.
 PR 25-MAY-2000; 2000US-0207454P.
 PR 09-JUN-2000; 2000US-021314P.
 PR 18-JUL-2000; 2000US-021907P.
 PR 13-DEC-2000; 2000US-0255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JB;
 XX
 DR MPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX
 PS Claim 1; Page 3785; 11750bp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)

CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX

SO Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;

Query Match 69.4%; Score 1810.4; DB 5; Length 2085;

Best Local Similarity 98.8%; Pred. No. 0; Mismatches 11; Indels 13; Gaps 13;

Matches 1960; Conservative 0; Indels 13; Gaps 13;

QY 3 CGGTGGCCGTCCGCCCACTCCGCCGCTTCGGGGAAATGGCTGCGAGACCTTAGAGGCT 62
DB 17 CGGTGGCCGTCCGCCCACTCCGCCGCTTCGGGGAAATGGCTGCGAGACCTTAGAGGCT 76
QY 63 GCGGAGCTTACTCAGCGGAAACAGCTCTAGATATCTGAGTTGTTGAAATAGAAAGCC 122
DB 77 GCGGAGCTTACTCAGCGGAAACAGCTCTAGATATCTGAGTTGTTGAAATAGAAAGCC 135
QY 123 TGTACTGTGAAACAGTGGCTGACAGACAGTGTGTGTGAGCTGTGCTGTGAGAC 182
DB 136 TGTACTGTGAAACAGTGGCTGACAGACAGTGTGTGTGAGCTGTGCTGTGAGAC 195
QY 183 CAGAGGTTTCTGTGCAAGGCTTGTGTGATTTAGATTTCAAGGAAAGTGTCCA 242
DB 196 CAGAGGTTTCTGTGCAAGGCTTGTGTGATTTAGATTTCAAGGAAAGTGTCCA 254
QY 243 AGCTTTCAAGTTTGGAGAGGTATGAGACAGAAAGGCAACCGAGCAATGAGAGGCAAC 302
DB 255 AGCTTTCAAGTTTGGAGAGGTATGAGACAGAAAGGCAACCGAGCAATGAGAGGCAAC 313
QY 303 TAAAGGCAATCAAAACCAACAGCAAAAGTTTCAAGAAACATGAGGCTTTTCAGAGCAAC 362
DB 314 TAAAGGCAATCAAAACCAACAGCAAAAGTTTCAAGAAACATGAGGCTTTTCAGAGCAAC 373
QY 363 TATTCGCAAGGAGAGGCGCGAGGAGACGCGAGGCTTCACTGAGGCTCAACCCCTC 422
DB 374 TATTCGCAAGGAGAGGCGCGAGGAGACGCGAGGCTTCACTGAGGCTCAACCCCTC 432
QY 423 ACAGACAGAGTGGGCTGTCTCTGCGCGAGTGGAGAGGAGCCCAAGGCGCACTGAGAG 482
DB 433 ACAGACAGAGTGGGCTGTCTCTGCGCGAGTGGAGAGGAGCCCAAGGCGCACTGAGAG 492
QY 483 CGTGGAGCAAGTTCTGACCAATTTGCGCGCGCGCGAGAGAGCAATGCTGTCTCCCT 542
DB 493 CGTGGAGCAAGTTCTGACCAATTTGCGCGCGCGCGAGAGAGCAATGCTGTCTCCCT 552
QY 543 GAGAGATTTCTGTGAGCCCAAGTCTCTGCGCGCGCAAGAGCGCGAGACGCTTCGAGAG 602
DB 553 GAGAGATTTCTGTGAGCCCAAGTCTCTGCGCGCGCAAGAGCGCGAGACGCTTCGAGAG 610
QY 603 CAGCTGTGAAAGCGCTTCTGAGACCAAGAGGCGCCCAAGTCTCTCTCAAGCTGTGAA 662
DB 611 CAGCTGTGAAAGCGCTTCTGAGACCAAGAGGCGCGCCCAAGTCTCTCTCAAGCTGTGAA 668
QY 663 GGAACGACCAAGCTCTTCTGAAAGAGTGAAGAGGAGATGACAGATGACACTCCCA 722
DB 669 GGAACGACCAAGCTCTTCTGAAAGAGTGAAGAGGAGATGACAGATGACACTCCCA 727
QY 723 TAGTGAACGATGAGCTTGAAGAGCTTGAAGAGCTTGAAGAGCTTGAAGAGCTTGA 782
DB 728 TAGTGAACGATGAGCTTGAAGAGCTTGAAGAGCTTGAAGAGCTTGAAGAGCTTGA 785
QY 783 ACAGAGACCCCACTGAGAGGCGCGTGAAGAGATGAGAGGCTGTGCGAGAGAGGCGCG 842
DB 786 ACAGAGACCCCACTGAGAGGCGCGTGAAGAGATGAGAGGCTGTGCGAGAGAGGCGCG 845
QY 843 GGAAGAGGCTTCCCGCGAGACTGTGTGAGCTCGAGGCGAGTGAACATGTGTGAGAGGCT 902
DB 846 GGAAGAGGCTTCCCGCGAGACTGTGTGAGCTCGAGGCGAGTGAACATGTGTGAGAGGCT 905
QY 903 GCCCAGTAAAGAGAGCCCGAGAAAGATCAAGAGGCTGTGTCCAGGCTGTGAGAAAGATGA 962
DB 906 GCCCAGTAAAGAGAGCCCGAGAAAGATCAAGAGGCTGTGTCCAGGCTGTGAGAAAGATGA 965

QY 963 CAGAGAGTAAAGTGGAGGAAAGGCGCTCAGAGATCAAGATGAGAGGCTGAGAG 1022
DB 966 CAGAGAGTAAAGTGGAGGAAAGGCGCTCAGAGATCAAGATGAGAGGCTGAGAG 1025
QY 1023 CTTGGGCTGACCGAAGCTTGAATGTGAGGTTACAGCCCAAGCCCTGTATTTGATTTG 1082
DB 1026 CTTGGGCTGACCGAAGCTTGAATGTGAGGTTACAGCCCAAGCCCTGTATTTGATTTG 1085
QY 1083 CCGGCACTTCAACAAACAGGTTTAAATTTGTGTGAGCCGTGTGAAAGATGATTTCA 1142
DB 1086 CCGGCACTTCAACAAACAGGTTTAAATTTGTGTGAGCCGTGTGAAAGATGATTTCA 1145
QY 1143 TGGCGATTGTGGGCTATTTCTGAGGCTGAGGAGGCTTTTGAAGAAAGTGGAGAGA 1202
DB 1146 TGGCGATTGTGTGGCATTTCTGAGGCTGAGGAGGCTTTTGAAGAAAGTGGAGAGA 1205
QY 1203 CTATATCTGCGCAACCTGACCAATTTCTGCAAGTGAAGATGACATCTTCAAGAAAGCG 1262
DB 1206 CTATATCTGCGCAACCTGACCAATTTCTGCAAGTGAAGATGACATCTTCAAGAAAGCG 1265
QY 1263 AGATCAGAGAGAGCTTAAATGAGACCTGAGATGCTGATGAGCAACCGATTGACAGAT 1322
DB 1266 AGATCAGAGAGAGCTTAAATGAGACCTGAGATGCTGATGAGCAACCGATTGACAGAT 1325
QY 1323 AGAACAATGAGAGAGAGCTTACCGAAGACCAAGGATTAAGGATTAAGATTAAGAAAGC 1382
DB 1326 AGAACAATGAGAGAGAGCTTACCGAAGACCAAGGATTAAGGATTAAGATTAAGAAAGC 1385
QY 1383 TGCATATTCAGATGGGCAAGAAACCTCAGATCTTCAAGCTGTGATGAGAGGCTGAG 1442
DB 1386 TGCATATTCAGATGGGCAAGAAACCTCAGATCTTCAAGCTGTGATGAGAGGCTGAG 1445
QY 1443 TGCCTCAAAATGATTTGGGCGCGGCTGTCAAGTGGCGAGCCGACTGCTGTACTG 1502
DB 1446 TGCCTCAAAATGATTTGGGCGCGGCTGTCAAGTGGCGAGCCGACTGCTGTACTG 1505
QY 1503 CAGTAAATGACTGTATCTCAACACAGCGCGAGCAACATGAAGTTTCTAAGCTCAGTAA 1562
DB 1506 CAGTAAATGACTGTATCTCAACACAGCGCGAGCAACATGAAGTTTCTAAGCTCAGTAA 1565
QY 1563 AGAA CAGAGACCAAGCTTAAAGAAAGATGAAGTGAAGCCAGAGAGCCCAAGCTTCC 1622
DB 1566 AGAA CAGAGACCAAGCTTAAAGAAAGATGAAGTGAAGCCAGAGAGCCCAAGCTTCC 1625
QY 1623 GAAATGCGGTCTCAGAGAGGATTTAAATCTCTTGTGTGACAAAGAGCAGCTCAGA 1682
DB 1626 GAAATGCGGTCTCAGAGAGGATTTAAATCTCTTGTGTGACAAAGAGCAGCTCAGA 1685
QY 1683 AAAAAAGAGACCAAGTGAAGAGAGAGTGTGTCTCTGCGGAGTGAAGCACTCGG 1742
DB 1686 AAAAAAGAGACCAAGTGAAGAGAGAGTGTGTCTCTGCGGAGTGAAGCACTCGG 1745
QY 1743 GAAAGAGACCTTGTGAGAGACAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1802
DB 1746 GAAAGAGACCTTGTGAGAGACAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1805
QY 1803 AGTAAAGCAAAAGAGT 1862
DB 1806 AGTAAAGCAAAAGAGT 1865
QY 1863 CTTAGAGGCTTGGCTCTGAGACCCCTCCGTTCTTCTGTGATGACATCTCCCTGGGCTG 1922
DB 1866 CTTAGAGGCTTGGCTCTGAGACCCCTCCGTTCTTCTGTGATGACATCTCCCTGGGCTG 1924
QY 1923 TCCAGAGCTGAGAGTGTGAGCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1982
DB 1925 TCCAGAGCTGAGAGTGTGAGC-TTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1983
QY 1983 GCGG 1986
DB 1984 GCGG 1987

RESULT 6

ABV22192 standard; cDNA, 2085 BP.

ABV22192;

13-SEP-2002 (first entry)

Human prostate expression marker cDNA 22183.

Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker; pharmacogenomic marker; gene; ss.

Homo sapiens.

W0200160860-A2.

23-AUG-2001.

20-FEB-2001; 2001WO-US005171.

17-FEB-2000; 2000US-0183319P.

16-MAR-2000; 2000US-0189862P.

25-MAY-2000; 2000US-0207454P.

09-JUN-2000; 2000US-0211314P.

18-JUL-2000; 2000US-0239007P.

13-DEC-2000; 2000US-0255281P.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Schlegel R, Endege WO, Monahan JB;

MPI; 2001-662795/76.

Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

Claim 1; Page 3816; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker

Sequence 2085 BP; 520 A; 526 C; 656 G; 378 T; 0 U; 5 Other;

Query Match 69.3%; Score 1808.8; DB 5; Length 2085;

Best Local Similarity 98.7%; Pred. No. 0; Mismatches 12; Indels 13; Gaps 13;

Matches 1959; Conservative 0;

3 CGGTGGCCGTCGCCCACTCCGCGGCTTCGGGGAATGCTGCGAGACCTTAGAGGCT 62

17 CGGTGGCCGTCGCCCACTCCGCGGCTTCGGGGAATGCTGCGAGACCTTAGAGGCT 76

63 CGGAGCTTACTCCAGCGGGAACGCTTATGATATCTGAGTGTGGAATAACAAAGC 122

77 GCGGAGCTTACTCCAGCGGGAACGCTTATGATATCTGAG- TGTGAAAAATACAAAGC 135

123 TGTACTCTGTGAACAGTGGCTGACAAAGTGTGTGTGAGGCTGGCTGTCTGTGAGC 182

136 TGTACTCTGTGAACAGTGGCTGACAAAGTGTGTGTGAGGCTGGCTGTCTGTGAGC 195

183 CCAAGAGTTTGTCTGCGAGGGTTTTGTGTATTTAGATTTCAGGAGAAAAGTGTCCA 242

DB 196 ||||| 254

CCAGAGGTTTGTCTGCGAGGG- TTTGTGTATTTAGATTTCAGGAGAAAAGTGTCCA 254

QY 243 AGCTTCAAGTGTGAGAGGATATGACGACAAAGGCGCACCGAGCAATGAGAGCAC 302

DB 255 AGCTTCAAGTGTGAGAGGATATGACGACAAAGGCGCACCGAGCAATGAGAGCAC 313

QY 303 TAAAGCCATCAAAACCACCAAGCAAAAGTTCAAGAAAATATGGGTTTTGAAAGCAC 362

DB 314 TAAAGCCATCAAAACCACCAAGCAAAAGTTCAAGAAAATATGGGTTTTGAAAGCAC 373

QY 363 TATGCCAAGGAGAGGCGCAGGAGGCTGACCCACTGAGAGGCGCACCGCC 422

DB 374 TATGCCAAGGAGAGGCGCAGGAGGCTGACCCACTGAGAGGCGCACCGCC 432

QY 423 ACAGCAGAGCTGGGCTGTCTCCCTGCGGCAAGTGGAGGCGCACCGCC 482

DB 433 ACAGCAGAGCTGGGCTGTCTCCCTGCGGCAAGTGGAGGCGCACCGCC 492

QY 483 CGTGAAGAGTTCTGACCAATTCGCGCGCGCGCGCAGAGAGCATGCTGTCTCCT 542

DB 493 CGTGAAGAGTTCTGACCAATTCGCGCGCGCGCGCAGAGAGCATGCTGTCTCCT 552

QY 543 GGAAGATTTCTGTGAGGCGCCACGTCTGCGCGCGCGCACAGGCGCACCGCC 602

DB 553 GGAAGATTTCTGTGAGGCGCCACGTCTGCGCGCGCACAGGCGCACCGCC 610

QY 603 CAGGTGGAAGCGCTTCTGAGACCAAGAGCGGCGCCAGCTGTCTGCAAGTGTGA 662

DB 611 CAGGTGGAAGCGCTTCTGAGACCAAGAGCGGCGCCAGCTGTCTGCAAG- TGTGA 668

QY 663 GGAACGACGAGCTTCTTGAAGGTGAAGAGGAGATGACACGATGACACTTCCGA 722

DB 669 GGAACGACGAG- CTCTTCTGAAGAGGTGAAGAGGAGATGACACGATGACACTTCCGA 727

QY 723 TAGTCAAGGATGAGCTGACCTTGAAGAGCTTCAAGATTCGCGGAGAGGAGGAG 782

DB 728 TAGTCAAGGATGAGCTGACCTTGAAGAGCTTCAAGATTCGCGGAGAGGAGGAG 785

QY 783 ACAGAGCCCACTGAGAGGCGCCCTGAAGAGGATGACAGTCTGCGGAGAGAGCGCG 842

DB 786 ACAGAGCCCACTGAGAGGCGCCCTGAAGAGGATGACAGTCTGCGGAGAGAGCGCG 845

QY 843 GGAAGAGGTTCCGCGAGACTGTGAGCTTCCAGGCGCACTGTGAGAGGCGCTCT 902

DB 846 GGAAGAGGTTCCGCGAGACTGTGAGCTTCCAGGCGCACTGTGAGAGGCGCTCT 905

QY 903 GCGCAGTAAAGCAGAGGCGCCGAGAGATGACAGTGGGCTGTGCTCCAGGCTGGAGAAAGATGA 962

DB 906 GCGCAGTAAAGCAGAGGCGCCGAGAGATGACAGTGGGCTGTGCTCCAGGCTGGAGAAAGATGA 965

QY 963 CAGAGAGGTAAGTGAAGGGAAGGCGGCTCAGGACATCAAGATGAGAGGCTTGA 1022

DB 966 CAGAGAGGTAAGTGAAGGGAAGGCGGCTCAGGACATCAAGATGAGAGGCTTGA 1025

QY 1023 CTTGGGCGGACCGAAGCTTGAATGTGAGGTTACAGCCCAAGCTTGTATTCATTTG 1082

DB 1026 CTTGGGCGGACCGAAGCTTGAATGTGAGGTTACAGCCCAAGCTTGTATTCATTTG 1085

QY 1083 CCGCAGCTTCAACAAACAGTATATGATTTGCTGTGACCGCTGTGAAGAAATGTTTCA 1142

DB 1086 CCGCAGCTTCAACAAACAGTATATGATTTGCTGTGACCGCTGTGAAGAAATGTTTCA 1145

QY 1143 TGGGATTTGTGAGGCAATTTCTGAGGCTGAGAGGCGCTTTTGAAGAAATGAGGAGA 1202

DB 1146 TGGGATTTGTGAGGCAATTTCTGAGGCTGAGAGGCGCTTTTGAAGAAATGAGGAGA 1205

QY 1203 CTATATCTGCCCAACAGCAATTCAGAGTGAAGATGAGATCTCATTCAGAAAGCGC 1262

DB 1206 CTATATCTGCCCAACAGCAATTCAGAGTGAAGATGAGATCTCATTCAGAAAGCGC 1265

QY 1263 AGATGAGGAGAGGATTAATGAGACTGTGAGATCTGATGCAAGATTTGATCAAGATTT 1322

Db	1266	AGATCAGCAGGAACCTAAATGGAGACTGGAGATGCTGATGGCACCCGATTGTACAAATAT	1325
Qy	1323	AGGAACAATAGAGCAAGAAATCTAGCCGAAGACCAAGGATTAAGGGTGAATTTGAAAGC	1382
Db	1326	AGGAACAATAGAGCAGAAATCTAGCCGAAGACCAAGGGTAAAGGGTGAATTTGAAAGC	1385
Qy	1383	TGCAAATTCGAATGAGCAAGAAAGAAATCTCAATCTTCACAGCTGTGATAGAGGCGCTGG	1442
Db	1386	TGCAAAATTCGAATGAGCAAGAAAGAAATCTCAATCTTCACAGCTGTGATAGAGGCGCTGG	1445
Qy	1443	TGCCTCAAAATGTATTTGGCCCCGGGTGCTGCAGTGGACGACGCCCATCTCGTGTACTG	1502
Db	1446	TGCCTCAAAATGTATTTGGCCCCGGGTGCTGCAGTGGACGACGCCCATCTCGTGTACTG	1505
Qy	1503	CAGTAAATGACTGTATCTCTCAACAACGCGCGCAGCAATGAAGTTTCTAAGCTCAGGTAA	1562
Db	1506	CAGTAAATGACTGTATCTCTCAACAACGCGCGCAGCAATGAAGTTTCTAAGCTCAGGTAA	1565
Qy	1563	AGAAACAGAACCCAAAGCCTTAAGAAAGATGAAGTGAAGCCAGAGAACCCATCTTCC	1622
Db	1566	AGAAACAGAACCCAAAGCCTTAAGAAAGATGAAGTGAAGCCAGAGAACCCATCTTCC	1625
Qy	1623	GAAATGGCGTCTCAGAGCAGGTATTTAAATCTCTCTGTGCAACAAGAACCCAGCTCCAGA	1682
Db	1626	GAAATGGCGTCTCAGAGCAGGTATTTAAATCTCTCTGTGCAACAAGAACCCAGCTCCAGA	1685
Qy	1683	AAAAAAGAGACCAACAGTAAAGAAAGGCAGTGGTCCCTCGCGGAGTGAAGCACTCGG	1742
Db	1686	AAAAAAGAGACCAACAGTAAAGAAAGGCAGTGGTCCCTCGCGGAGTGAAGCACTCGG	1745
Qy	1743	GAAGGAGCAGCTTGTGTGAGAGCAGCAGCCGTCGTGGGCGAGCGATCACAATTAATGC	1802
Db	1746	GAAGGAGCAGCTTGTGTGAGAGCAGCAGCCGTCGTGGGCGAGCGATCACAATTAATGC	1805
Qy	1803	AGTAAAGCCAAATAAGACTGCGTCCCTCGCGCGCTCACTGTTGATTAATGTAATGATCA	1862
Db	1806	AGTAAAGCCAAATAAGACTGCGTCCCTCGCGCGCTCACTGTTGATTAATGTAATGATCA	1865
Qy	1863	CCTAGGGGTTGGCCTCTGTGACCCCTCCCGTCTTTCTGATAGCAATCCCTGGGCGTG	1922
Db	1866	CCTAGGGGTTGGCCTCTGTGACCCCTCCCGTCTTTCTGATAGCAATCCCTGGGCGTG	1924
Qy	1923	TTCAGGACTGGAGAGTTGAGCTTTGTGTTAAGCTGATCAACAGACACCGGCTGCACATCA	1982
Db	1925	TTCAGGACTGGAGAGTTGAGCAGCCTTGTGTTAAGCTGATCAACAGACACCGGCTGCAGGA	1983
Qy	1983	GCAG 1986	
Db	1984	GCAG 1987	

RESULT 7	
AACT7933	
ID	AACT7933 standard; cDNA, 1764 BP.
XX	
XX	
AC	
XX	AACT7933;
XX	
DT	08-FEB-2001 (first entry)
DE	
XX	Human cancer associated gene sequence SEQ ID NO:327.
KM	Human; cancer associated gene; cancer antigen; detection; cancer;
KM	diagnosis; cytostatic; proliferative; vunerary; immunomodulator;
KM	antidiabetic; antisthmatic; antineumatic; antiarthritis; antiviral;
KM	antiinflammatory; antihypoid; antiallergic; antibacterial; cardiant;
KM	dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KM	vasotrophic; antiproliferic; antiangiogenic; gene therapy; inflammation;
KM	immune disorder; haematopoietic cell disorder; autoimmune disorder;
KM	allergic reaction; graft versus host disease; organ rejection;
KM	haemostatic; thrombolytic; cardiovascular disorder; infection;
XX	neurological disease; drug screening; ss.
OS	Homo sapiens.

XX	MO200055350-AL.
PN	
XX	21-SEP-2000.
PD	
XX	08-MAR-2000; 2000MO-US005882.
PF	
XX	12-MAR-1999; 99US-0124270P.
PR	
XX	(HUMA-) HUMAN GENOME SCI INC.
XX	
PA	
XX	
PI	Rosen CA, Ruben SM;
XX	
XX	WPI; 2000-587533/55.
DR	P-PSDB; AAB43724.
XX	
XX	
PT	Novel isolated nucleic acids comprising sequences encoding peptides useful for treating or diagnosing e.g. cancer.

Claim 1; Page 880-881; 2352pp; English.

AACT7607 to AAC78444 encode the human cancer associated proteins given in AAB33398 to AAB44233. The proteins can have activities based on the tissues and cells the genes are expressed in. Example of activities include: cytostatic; proliferative; vulnerability; immunomodulator; anticarcinogenic; antiautomatic; antirheumatic; antithrombotic; antidiabetic; antihypertensive; antiallergic; antibacterial; antiviral; antineoplastic; neuroprotective; cardiac; thrombolytic; coagulant; dematological; vasotropic; antipsoriatic and angiogenic. The polypeptides and polypeptides can be used for preventing, treating or ameliorating medical conditions and diagnosing pathological conditions. The polypeptides, polypeptides, antibodies, agonists and antagonists from the present invention may be used to treat immune disorders by activating or inhibiting the proliferation, differentiation or mobilisation of immune cells, to treat disorders of haematopoietic cells, autoimmune disorders, allergic reactions, graft versus host disease and organ rejection, modulate haemostatic or thrombolytic activity, modulate inflammation, cancers, cardiovascular disorders, neurological disease and bacterial or viral infections. The peptides, nucleotides, antibodies, agonists and antagonists may be also be used in drug screens. AAC78445 to AAC78457 and AAB44240 represent sequences used in the exemplification of the present invention.

Sequence 1764 BP: 468 A; 424 C; 415 G; 453 T; 0 U; 4 Other;

59.9%; Score 1564.4; DB 3; Length 1764;

Best Local Similarity 99.6%; Pred.No. 0;
Matches 1610; Conservative 0; Mismatches 2; Indels 5; Gaps 4;

996 CGACATCAAGATGAGGAGCCTGAGAC-TTGGGCCGACCGAAGCCTGAATGTGAGGCTT 1054

1 GGAACATCAAGATGAGGAGCGCTGCTTCTGCTGACCGAGAGCGCTGATGTGAGCGCTT 60

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

1055
1060
1065
1070
1075
1080
1085
1090
1095
1100
1105
1110
1115
1120
1125
1130
1135
1140
1145
1150
1155
1160
1165
1170
1175
1180
1185
1190
1195
1200
1205
1210
1215
1220
1225
1230
1235
1240
1245
1250
1255
1260
1265
1270
1275
1280
1285
1290
1295
1300
1305
1310
1315
1320
1325
1330
1335
1340
1345
1350
1355
1360
1365
1370
1375
1380
1385
1390
1395
1400
1405
1410
1415
1420
1425
1430
1435
1440
1445
1450
1455
1460
1465
1470
1475
1480
1485
1490
1495
1500
1505
1510
1515
1520
1525
1530
1535
1540
1545
1550
1555
1560
1565
1570
1575
1580
1585
1590
1595
1600
1605
1610
1615
1620
1625
1630
1635
1640
1645
1650
1655
1660
1665
1670
1675
1680
1685
1690
1695
1700
1705
1710
1715
1720
1725
1730
1735
1740
1745
1750
1755
1760
1765
1770
1775
1780
1785
1790
1795
1800
1805
1810
1815
1820
1825
1830
1835
1840
1845
1850
1855
1860
1865
1870
1875
1880
1885
1890
1895
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
2030
2035
2040
2045
2050
2055
2060
2065
2070
2075
2080
2085
2090
2095
2100
2105
2110
2115
2120
2125
2130
2135
2140
2145
2150
2155
2160
2165
2170
2175
2180
2185
2190
2195
2200
2205
2210
2215
2220
2225
2230
2235
2240
2245
2250
2255
2260
2265
2270
2275
2280
2285
2290
2295
2300
2305
2310
2315
2320
2325
2330
2335
2340
2345
2350
2355
2360
2365
2370
2375
2380
2385
2390
2395
2400
2405
2410
2415
2420
2425
2430
2435
2440
2445
2450
2455
2460
2465
2470
2475
2480
2485
2490
2495
2500
2505
2510
2515
2520
2525
2530
2535
2540
2545
2550
2555
2560
2565
2570
2575
2580
2585
2590
2595
2600
2605
2610
2615
2620
2625
2630
2635
2640
2645
2650
2655
2660
2665
2670
2675
2680
2685
2690
2695
2700
2705
2710
2715
2720
2725
2730
2735
2740
2745
2750
2755
2760
2765
2770
2775
2780
2785
2790
2795
2800
2805
2810
2815
2820
2825
2830
2835
2840
2845
2850
2855
2860
2865
2870
2875
2880
2885
2890
2895
2900
2905
2910
2915
2920
2925
2930
2935
2940
2945
2950
2955
2960
2965
2970
2975
2980
2985
2990
2995
3000
3005
3010
3015
3020
3025
3030
3035
3040
3045
3050
3055
3060
3065
3070
3075
3080
3085
3090
3095
3100
3105
3110
3115
3120
3125
3130
3135
3140
3145
3150
3155
3160
3165
3170
3175
3180
3185
3190
3195
3200
3205
3210
3215
3220
3225
3230
3235
3240
3245
3250
3255
3260
3265
3270
3275
3280
3285
3290
3295
3300
3305
3310
3315
3320
3325
3330
3335
3340
3345
3350
3355
3360
3365
3370
3375
3380
3385
3390
3395
3400
3405
3410
3415
3420
3425
3430
3435
3440
3445
3450
3455
3460
3465
3470
3475
3480
3485
3490
3495
3500
3505
3510
3515
3520
3525
3530
3535
3540
3545
3550
3555
3560
3565
3570
3575
3580
3585
3590
3595
3600
3605
3610
3615
3620
3625
3630
3635
3640
3645
3650
3655
3660
3665
3670
3675
3680
3685
3690
3695
3700
3705
3710
3715
3720
3725
3730
3735
3740
3745
3750
3755
3760
3765
3770
3775
3780
3785
3790
3795
3800
3805
3810
3815
3820
3825
3830
3835
3840
3845
3850
3855
3860
3865
3870
3875
3880
3885
3890
3895
3900
3905
3910
3915
3920
3925
3930
3935
3940
3945
3950
3955
3960
3965
3970
3975
3980
3985
3990
3995
4000
4005
4010
4015
4020
4025
4030
4035
4040
4045
4050
4055
4060
4065
4070
4075
4080
4085
4090
4095
4100
4105
4110
4115
4120
4125
4130
4135
4140
4145
4150
4155
4160
4165
4170
4175
4180
4185
4190
4195
4200
4205
4210
4215
4220
4225
4230
4235
4240
4245
4250
4255
4260
4265
4270
4275
4280
4285
4290
4295
4300
4305
4310
4315
4320
4325
4330
4335
4340
4345
4350
4355
4360
4365
4370
4375
4380
4385
4390
4395
4400
4405
4410
4415
4420
4425
4430
4435
4440
4445
4450
4455
4460
44

61 ACGACCCCAAGCCCTGTATGATTTCGGCAGCCCTACACAACAGTTAAGATT 120

1115 GCTGTGACCGCTGTGAGATGGTTTCATGCGCATTGTGTGGGCATTTCTGAGGCTCGAG 1174

121 GCTGTACCGCTGTGAAGAATGTTTCATGGCGATTGTGTGGCATTCTGAGGCTCGAG 180

1175 GGAGGCTTTTGAAAGGAATGGGGAAGACTATATCTGCCCAACTGCACATTCTGCAAG 1234

181 GGAGGCTTTGGAAAGCAATGGGAGACTATATCTGCCAACTGCACCATTCGCAAG 240

1235 TGGAGGATGACACTCATTCAGAAACATCAGATCAGAGGAAAGCTAAATGAGACCTGAG 1294

[illegible]

241 TCGAGGATGGACTCATTCTGGTCCGCAGATCAGACGGGAACTTTATTAAGAACCCAAAA 1356

1295 ATGCTGATGGCACCAGATTGTACAGATATAGGAACATAGAGCAGAGATCTAGCGGAGACC 1354

301 ATGCTGATGGACCGATTGTACAACTATAGCAACATAGAGCAGAAGCTTAGCGAGACC 360

QY 1355 AAGGATTAAGGTTGATTAAGTGAAGAGCTGCATAATCCAACTGGCAAGAGAACTCAAGA 1414
DB 361 AAGGATTAAGGTTGATTAAGTGAAGAGCTGCATAATCCAACTGGCAAGAGAACTCAAGA 420
QY 1415 TCTTCCAGCCTGTGATTAAGGCGCTGGTCTCAAAAATGATTTGGCCCCGGGTCTGTG 1474
DB 421 TCTTCCAGCCTGTGATTAAGGCGCTGGTCTCAAAAATGATTTGGCCCCGGGTCTGTG 480
QY 1475 ACGTGGCGACGCCGATCTGGGTGATCTGAGTAAATGATCTGATCTCTCAACACGCGGAG 1534
DB 481 ACGTGGCGCA-CCGATCTGGGTGATCTGAGTAAATGATCTGATCTCTCAACACGCGGAG 539
QY 1535 CGACATTAAGTTTCTAAGCTCAAGTAAAGACAGAACCCAAAGCTTAAAGAAAGATGA 1594
DB 540 CGACATTAAGTTTCTAAGCTCAAGTAAAGACAGAACCCAAAGCTTAAAGAAAGATGA 599
QY 1595 AGATGAAGCCAGAGAGCCGATCTTCCGAAATGGCGGTCTCAAGCAGGTATTAATCT 1654
DB 600 AGATGAAGCCAGAGAGAGCCGATCTTCCGAAATGGCGGTCTCAAGCAGGTATTAATCT 659
QY 1655 CTTCGTGCAACAAGAGACAGCTCCAGAAAAAGAGACCAAGTGAAGAGCAGTGG 1714
DB 660 CTTCGTGCAACAAGAGACAGCTCCAGAAAAAGAGACCAAGTGAAGAGCAGTGG 719
QY 1715 TGGTCCCTGCGCGAGTGAAGCACTCGGGAAGAAAGACAGCTTGTGAAGACAGCAGCTG 1774
DB 720 TGGTCCCTGCGCGAGTGAAGCACTCGGGAAGAAAGACAGCTTGTGAAGACAGCAGCTG 779
QY 1775 CGTGGCGAGAGATGCAATTAATGCAATGCAATTAATGCAATTAATGCAATTAATGCA 1834
DB 780 CGTGGCGAGAGATGCAATTAATGCAATGCAATTAATGCAATTAATGCAATTAATGCA 839
QY 1835 CGTCACTGTGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATG 1894
DB 840 CGTCACTGTGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATG 899
QY 1895 CTTCGTGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATG 1954
DB 900 CTTCGTGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATG 959
QY 1955 CTGATCAAGACACCGGCTGACCAATCAAGCGGAAAGACAGCCCAATTCAGAGATGCC 2014
DB 960 CTGATCAAGACACCGGCTGACCAATCAAGCGGAAAGACAGCCCAATTCAGAGATGCC 2019
QY 2015 CTGCGGCCCTGTGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 2074
DB 1020 CTGCGGCCCTGTGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 2079
QY 2075 TAAACCTGACTGTAAGCGTTCACTTAATGATTAATGATTAATGATTAATGATTAATG 2134
DB 1080 TAAACCTGACTGTAAGCGTTCACTTAATGATTAATGATTAATGATTAATGATTAATG 2139
QY 2135 CAGGCTCTGTAAGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 2194
DB 1140 CAGGCTCTGTAAGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 2199
QY 2195 CGGCAAGCCACATCCACCCCTGTCTGACATGAGTTGTTCTGACAAAGCGCTGATAG 2254
DB 1200 CGGCAAGCCACATCCACCCCTGTGTGACATGAGTTGTTCTGACAAAGCGCTGATAG 2259
QY 2255 CTTCAGTTTTCACATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 2314
DB 1260 CTTCAGTTTTCACATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 2319
QY 2315 GTGGGAATCTTTCAGATTAAGTTGATTAATGATTAATGATTAATGATTAATGATTAAT 2374
DB 1320 GTGGGAATCTTTCAGATTAAGTTGATTAATGATTAATGATTAATGATTAATGATTAAT 2379
QY 2375 GGGGAGATAGGGGATTTTTCATGATGATTCAGATTCAGATTCAGATTCAGATTCAGAT 2434
DB 1380 GGGGAGATAGGGGATTTTTCATGATGATTCAGATTCAGATTCAGATTCAGATTCAGAT 2437
QY 2435 AACACATTCGAATTCGAAGG-CACACACTCTGCTTCATTAAGGCCCAAGGTAAGT 2493

DB 1438 AACACATTCGAATTCGAAGG-CACACACTCTGCTTCATTAAGGCCCAAGGTAAGT 1497
QY 2494 TCACACTAGAACACTGTCTGACCGAGACGCGTGCCTTGAGACTTGATTAATTCATG 2553
DB 1498 TCACACTAGAACACTGTCTGACCGAGACGCGTGCCTTGAGACTTGATTAATTCATG 1557
QY 2554 TGAATGCTTTCTTCCCTCGTCTCTGAAATTTAAGACTCTTAAGATATATCCG 2610
DB 1558 TGAATGCTTTCTTCCCTCGTCTCTGAAATTTAAGACTCTTAAGATATATCCG 1614

RESULT 8
AAK53352
ID AAK53352 standard; cDNA; 7290 BP.
XX
AC AAK53352;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 2881.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN W0200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00466914.
XX
PR 27-APR-2000; 2000US-00560875.
XX
PR 20-JUN-2000; 2000US-00598075.
XX
PR 19-JUL-2000; 2000US-00620325.
XX
PR 01-SEP-2000; 2000US-00654936.
XX
PR 15-SEP-2000; 2000US-00663561.
XX
PR 20-OCT-2000; 2000US-00693325.
XX
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HVS-E) HVS-EQ INC.
XX
PI Tang YF, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao Qa, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR MPI; 2001-476283/51.
XX
DR P-PSDB; AAM80219.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 1; Page 5053-5055; 6221pp; English.
XX
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
XX (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
XX sequence listing were missing at the time of publication
XX
SQ Sequence 7290 BP; 1762 A; 1953 C; 1975 G; 1600 T; 0 U; 0 Other;

Query Match	49.1%	Score 1280.6;	DB 4;	Length 7290;
Best Local Similarity	92.6%;	Pred. No. 0;		
Matches 1401; Conservative	0;	Mismatches 4;	Indels 108;	Gaps 1;

QY	455	GTGGAGGACAGCCCAAGCGACATGAGGCGCTGGAGCAGTTCTCTGACCATTTGGCGGAGCGCC	514
Db	64	GTGGAGGACAGCCCAAGCGACCTGAGGCGCTGGAGCAGTTCTCTGACCATTTGGCGGAGCGCC	123
QY	515	GCGGCAGAGAGAGCATGCTCTGTCTCCCTCGAGAGATTCTGCTGAGACCCAGCTCTGCCCCG	574
Db	124	GCGGCAGAGAGAGCATGCTCTGTCTCTCCCTCGAGAGATTCTGCTGAGACCCAGCTCTGCCCCG	183
QY	575	CCACAGACGCCCGAGACACGCTCCGAGGGCAGCGCTGGAAAGCGCTTCTTGAGACCAAGAACCG	634
Db	184	CCACAGACGCCCGAGACACGCTCCGAGGGCAGCGCTGGAAAGCGCTTCTTGAGACCAAGAACCG	243
QY	635	GCCCCGAGCTGCTGCTCCACAGCTGTGAAAGAACAGACAGCCTCTTCTGAAAGAGTGAAAG	694
Db	244	GCCCCGAGCTGCTGCTCCACAGCTGTGAAAGAACAGACAGCCTCTTCTGAAAGAGTGAAAG	303
QY	695	GAGGGAGTGAACAAGATGACACTCTCCGATAGTGAACAGCGATGCGCTTGAACCTTGAAAGAC	754
Db	304	GAGGGAGTGAACAAGATGACACTCTCCGATAGTGAACAGCGATGCGCTTGAACCTTGAAAGAC	363
QY	755	TTCAGAATCCGCTTTCGCAAGAGGCGGGACAGAGGCCCATCTGAGAGGCCCTCTGAAGAGGA	814
Db	364	TTCAGAATCCGCTTTCGCAAGAGGCGGGACAGAGGCCCATCTGAGAGGCCCTCTGAAGAGGA	423
QY	815	TTCCAGATCTGCGCTTCGGAAAGAGGCGCGGAGAGAGGTCCCGCCGAGACTGTGCGGCTCCG	874
Db	424	TTCCAGATCTGCGCTTCGGAAAGAGGCGCGGAGAGAGGTCCCGCCGAGACTGTGCGGCTCCG	483
QY	875	AGGCCAGTGAACACTGTGAGGGGCGTCTCTGCCAGTAAAGCAGAGACCCGAGAACGATCAGG	934
Db	484	AGGCCAGTGAACACTGTGAGGGGCGTCTCTGCCAGTAAAGCAGAGACCCGAGAACGATCAGG	543
QY	935	GGCTTTGTGTCCCAAGGCTTCGGAAAGATATGACAGAGAGATTAAGTTGAGAGGAAAGCGCGCTC	994
Db	544	GGCTTTGTGTCCCAAGGCTTCGGAAAGATATGACAGAGAGATTAAGTTGAGAGGAAAGCGCGCTC	603
QY	995	AGGACATCTAAAGATGAGGAGCGCTGAGACCTTGAGGCCGACCGAAGCCTGAATGTGAGGTT	1054
Db	604	AGGACATCTAAAGATGAGGAGCGCTGAGACCTTGAGGCCGACCGAAGCCTGAATGTGAGGTT	663
QY	1055	ACGACCCCAAGCGCCCTGTATTGCAATTTGCGCGACAGCCTCACAAACAACAGTTTATGATTT	1114
Db	664	ACGACCCCAAGCGCCCTGTATTGCAATTTGCGCGCGACAGCCTCACAAACAACAGTTTATGATTT	723
QY	1115	GCTGTGACCGCTGTGMAAGAAATGTTTTCAATGGCCGATTGTGTGGGGAATTTCTGAGGCTGAG	1174
Db	724	GCTGTGACCGCTGTGMAAGAAATGTTTTCAATGGCCGATTGTGTGGGGAATTTCTGAGGCTGAG	783
QY	1175	GGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCCAATCTGCAACATTTCTGCAAG	1234
Db	784	GGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCCAATCTGCAACATTTCTGCAAG	843
QY	1235	TGCAGATGAGACTCATTTCAAGAAACGCGCAGATCAAGCAGAAAGCTTAATATGAGACTTGGAG	1294
Db	844	TGCAGATGAGACTCATTTCAAGAAACGCGCAGATCAAGCAGAAAGCTTAATATGAGACTTGGAG	903
QY	1295	ATGCTGATGGAACCGATTGTTCATCAAGTTATAGGAACATATGACACAAAGCTTAGCGAGAGCC	1354
Db	904	ATGCTGATGGAACCGATTGTTCATCAAGTTATAGGAACATATGACACAAAGCTTAGCGAGAGCC	963
QY	1355	AAGGGATPAAAGGTATGAAATTGAGAAAGCTGCAAAATCCAAATGTGGCAGAGAAAGAACTCAGA	1414
Db	964	AAGGGATPAAAGGTATGAAATTGAGAAAGCTGCAAAATCCAAATGTGGCAGAGAAAGAACTCAGA	1023
QY	1415	TCCTTCAGGCT-----	1425
Db	1024	TCCTTCAGGCTTCGATCCCGGCGCTATCCCAACAGAGCTGCTCTCTGCGAGGTATTTGG	1083

QY	1426	-----G 1426
Db	1084	AAATTCGTGTGTCTAGAAAGACTCTGCCTTCACTCTTGCACTGCAATTAGCTTCGAAG 1143
QY	1427	TGATPAGAGGCGCCCTGAGGCTCCAAATATTTGGCCCGGGTGTCTCACTGGCGCAGC 1486
Db	1144	TGATPAGAGGCGCCCTGAGGCTCCAAATATTTGGCCCGGGTGTCTCACTGGCGCAGC 1203
QY	1487	CCGACTCGGTGTACTGCAAGTATTAATCTGTATCTTCAACAACGCGCAGCGCAATATGAAT 1546
Db	1204	CCGACTCGGTGTACTGCAAGTATTAATCTGTATCTTCAACAACGCGCAGCGCAATATGAAT 1263
QY	1547	TTCTTAAGCTCAGTAAAGAACAGAAAGCCAAAGCTTAAAGAAAAGTGAAGTGAAGCCAG 1606
Db	1264	TTCTTAAGCTCAGTAAAGAACAGAAAGCCAAAGCTTAAAGAAAAGTGAAGTGAAGCCAG 1323
QY	1607	AGAAAGCCAGTCTTCCGAAATTCGCGTCTCAGGCAAGGATTTAAATCTCTTCTGTGCACA 1666
Db	1324	AGAAAGCCAGTCTTCCGAAATTCGCGTCTCAGGCAAGGATTTAAATCTCTTCTGTGCACA 1383
QY	1667	AGAAACCAAGCTCCAGAAAAAAAAGAGACAAGTGAAGAAAGCAGTGTGTCCCTGCGC 1726
Db	1384	AGAAACCAAGCTCCAGAAAAAAAAGAGACAAGTGAAGAAAGCAGTGTGTCCCTGCGC 1443
QY	1727	GGAAGTGAAGCACTCGGGAAGGAAGCACTTGTGAAGACAGACGCGCTGTGGCGAGCG 1786
Db	1444	GGAAGTGAAGCACTCGGGAAGGAAGCACTTGTGAAGACAGACGCGCTGTGGCGAGCG 1503
QY	1787	ATCAACAATTACAATGCAAGTAAAGCCAGAAAAAGACTGCTGCTCCCTCGCCGTACATGTTGT 1846
Db	1504	ATCAACAATTACAATGCAAGTAAAGCCAGAAAAAGACTGCTGCTCCCTCGCCGTACATGTTGT 1563
QY	1847	ATAAATGTATGTATA 1859
Db	1564	ATAAATCCACGAA 1576

RESULT 9	
AAK52368	
ID	AAK52368 standard; cDNA; 7838 BP.
XX	
AC	AAK52368;
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human polynucleotide SEQ ID NO 913.
XX	
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;
KW	nervous system disorder; arthritis; inflammation; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200157190-A2.
XX	
PD	09-AUG-2001.
XX	
PF	05-FEB-2001; 2001WO-US004098.
XX	
PR	03-FEB-2000; 2000US-00496914.
PR	27-APR-2000; 2000US-00560875.
PR	20-JUN-2000; 2000US-00598075.
PR	19-JUL-2000; 2000US-00620325.
PR	01-SEP-2000; 2000US-00634936.
PR	15-SEP-2000; 2000US-00633561.
PR	20-OCT-2000; 2000US-00693325.
PR	30-NOV-2000; 2000US-00728422.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX WPI: 2001-476283/51.
 DR P-PSDB: AAM79235.
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 PS Claim 1, Page 3007-3013; 6221pp; English.
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAW80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX
 SQ Sequence 7838 BP; 1872 A; 2123 C; 2118 G; 1725 T; 0 U; 0 Other;
 Query Match 49.1%; Score 1280.6; DB 4; Length 7838;
 Best Local Similarity 92.6%; Pred. No. 0;
 Matches 1401; Conservative 0; Mismatches 4; Indels 108; Gaps 1;
 455 GTGGAGGCAAGCCCAAGCGCACTGAGCGCGTGGAGCAATTCTGACCAATTGCGCGCCG 514
 599 GTGGAGGCAAGCCCAAGCGCACTGAGCGCGTGGAGCAATTCTGACCAATTGCGCGCCG 658
 515 GCGGCAAGAGAGAGCATGCT 574
 659 GCGGCAAGAGAGAGCATGCT 718
 515 CCACAGAGCGCGAG 634
 719 CCACAGAGCGCGAG 778
 635 GCCCCAGCTCTGCTTCAACAGCTGTGAAGGAACAGACAGCTCTTCTGAAAAGTGAAG 694
 779 GCCCCAGCTCTGCTTCAACAGCTGTGAAGGAACAGACAGCTCTTCTGAAAAGTGAAG 838
 695 GAGGGAGATGACAGATGACAGCTTCCGATGTGACAGAGAGAGAGAGAGAGAGAGAGAG 754
 839 GAGGGAGATGACAGATGACAGCTTCCGATGTGACAGAGAGAGAGAGAGAGAGAGAGAG 898
 755 TTCAAGATCGCTTGGAG 814
 899 TTCAAGATCGCTTGGAG 958
 815 TCCAGAGTCTGCTGCGAAG 874
 959 TCCAGAGTCTGCTGCGAAG 1018
 875 AGGCGATGACAGCTGTGAG 934
 1019 AGGCGATGACAGCTGTGAG 1078
 935 GGGTGTCTCCAGAGCTGGGAAAGATGACAGAGAGATTAAGTTGAGAGGAAAGGCGCTC 994
 1079 GGGTGTCTCCAGAGCTGGGAAAGATGACAGAGAGATTAAGTTGAGAGGAAAGGCGCTC 1138
 995 AGGACATCAAGAGATGAGAGCTTGAGAGCTTGAGAGCTTGAGAGCTTGAGAGCTTGAGAG 1054
 1139 AGGACATCAAGAGATGAGAGCTTGAGAGCTTGAGAGCTTGAGAGCTTGAGAGCTTGAGAG 1198
 1055 AGGACATCAAGAGCTTGAGAGCTTGAGAGCTTGAGAGCTTGAGAGCTTGAGAGCTTGAGAG 1114
 1199 AGGACATCAAGAGCTTGAGAGCTTGAGAGCTTGAGAGCTTGAGAGCTTGAGAGCTTGAGAG 1258

QY 1115 GCTGTACCGCTGTGAGAGATGTTTCATGCGGATTTGTGTGGCATTTCTGAGGCTCGAG 1174
 DB GCTGTACCGCTGTGAGAGATGTTTCATGCGGATTTGTGTGGCATTTCTGAGGCTCGAG 1318
 QY 1175 GAGGCTTTTGGAAAGAGATGGGAAAGATATATCTGCGCAATCTGACCATTTCTGCAAG 1234
 DB GAGGCTTTTGGAAAGAGATGGGAAAGATATATCTGCGCAATCTGACCATTTCTGCAAG 1378
 QY 1235 TCCAGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAGCTAAATGAGAGCTGAGAG 1294
 DB TCCAGATGAGACTCATTCAGAAACGGCAGATCAGCAGGAGCTAAATGAGAGCTGAGAG 1438
 QY 1295 ATGCTATGAGCAGCAGATTTGATACAGATATAGAAACATAGAGAGAGAGAGAGAG 1354
 DB ATGCTATGAGCAGCAGATTTGATACAGATATAGAAACATAGAGAGAGAGAGAGAGAG 1498
 QY 1355 AAGGATTAAGGATGAATTTGAGAAAGCTGCAATCCAGTGTGCAAGAAAGAACTCAAG 1414
 DB AAGGATTAAGGATGAATTTGAGAAAGCTGCAATCCAGTGTGCAAGAAAGAAAGCTCAAG 1558
 QY 1415 TCTTCAGCT----- 1425
 DB TCTTCAGCTGCTGCTCCGAGCTGTCTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1618
 QY 1426 -----G 1426
 DB AATTGCTGTGTCTGAGAGATCTGCGCTTCAACTGTGACATGATAGCTGAGAG 1678
 QY 1427 TGAATGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1486
 DB TGAATGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1738
 QY 1487 CCGACTCGGTGTACAGATATGATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1546
 DB CCGACTCGGTGTACAGATATGATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1798
 QY 1547 TTTCTAAGCTCAGGTAAAGACAGAAAGCCAAAGCTTAAAGAAAGATGAGAGAGAGAG 1606
 DB TTTCTAAGCTCAGGTAAAGACAGAAAGCCAAAGCTTAAAGAAAGATGAGAGAGAGAG 1858
 QY 1607 AGAAGCCAGCTTCTCGAAATCGGTGCTCAGGCAAGTATTAATCTCTGCTGACAG 1666
 DB AGAAGCCAGCTTCTCGAAATCGGTGCTCAGGCAAGTATTAATCTCTGCTGCTGACAG 1918
 QY 1667 AGAGACAGCTTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1726
 DB AGAGACAGCTTCAAGAAAG 1978
 QY 1919 AGAGACAGCTTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1978
 DB AGAGACAGCTTCAAGAAAG 2038
 QY 1727 GGAAGTGAAGCACTCGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1786
 DB GGAAGTGAAGCACTCGGAAAG 2038
 QY 1787 ATCACAATTACAGTCAAGTAAAGCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1846
 DB ATCACAATTACAGTCAAGTAAAGCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2098
 QY 2039 ATCACAATTACAGTCAAGTAAAGCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 2098
 DB ATCACAATTACAGTCAAGTAAAGCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2098
 QY 1847 ATTAATGTATGTA 1859
 DB ATTAATGTATGTA 2111
 DB ATTAATGTATGTA 2111
 RESULT 10
 AAS86576 standard; cDNA; 2332 BP.
 AAS86576;
 13-FEB-2002 (first entry)
 DNA encoding novel human diagnostic protein #22380.
 Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
OS
PN MO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSB-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX P-PSDB; ABG23389.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 22380; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostic as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS6197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2332 BP; 511 A; 621 C; 764 G; 436 T; 0 U; 0 Other;
Query Match 45.6%; Score 1190.8; DB 5; Length 2332;
Best Local Similarity 93.2%; Pred. No. 4.3e-310;
Matches 1362; Conservative 0; Mismatches 37; Indels 79; Gaps 7;
QY 244 GCTTCAGTGTGAGACAGGTATGACGACAAAGGCCACCGAGCATGAGAGGCACT 303
DB 370 GCTTACCTTACCGCTAGTATGACGACAAAGGCCACCGAGCATGAGAGGCACT 429
QY 304 AAGGCGATCAAAACCCACGAGGAAAGTTGAGAAAATAGGGGTTTTCAGAGGCACT 363
DB 430 AAGGCGATCAAAACCCACGAGGAAAGTTGAGAAAATAGGGGTTTTCAGAGGCACT 489
QY 364 ATCCCAAGCAGAGGCGCAGAGGAGCGGAGGCTGACCACTGAGCCGCCACCCCA 423
DB 490 ATCCCAAGCAGAGGCGCAGAGGAGCGGAGGCTGACCACTGAGCCGCCACCCCA 549
QY 424 CAGAGGAGCTGGGCTGTCTCTTGGGCGGCGAGTGGAGGCGAGCCCAAGCCGACTGAGGCG 483
DB 550 CAGAGGAGCTGGGCTGTCTCTTGGGCGGCGAGTGGAGGCGAGCCCAAGCCGACTGAGGCG 609
QY 484 GTGAGAGGCTTCGACCATTTGGCGGCGGCGGCGAGGAGGAGCATGCTGTCTCCCTG 543

DB 610 GTGAGAGGCTTCGACCATTTGGCGGCGGCGGCGAGGAGGAGCATGCTGTCTCCCTG 669
QY 544 GAGGATTCGTGTGAGGCCCAAGTCTCTGCCGCCGCAACAGCGCGAGAGCCCTCCGAGGCG 603
DB 670 GAGGATTCGTGTGAGGCCCAAGTCTCTGCCGCCGCAACAGCGCGAGAGCCCTCCGAGGCG 729
QY 604 AGCGTGAAGAGCGCTTTCAGAGACCAAGACGCGCCCAAGTCTGTCTTCACAGCTGTGAAG 663
DB 730 AGCGTGAAGAGCGCTTTCAGAGACCAAGAGCGCGCCCAAGTCTGTCTTCACAGCTGTGAAG 789
QY 664 GAACGACCAAGCTCTTTCGAAAAAGGTGAAGAGGAGATGACACAGATGACCTCCGAT 723
DB 790 GAACGACCAAGCTCTTTCGAAAAAGGTGAAGAGGAGATGACACAGATGACCTCCGAT 849
QY 724 AGTGAACGAGATGAGGCTGACCTTGAAGAGTTCGAAATGCGCTTTCGAGAGAGCGGAA 783
DB 850 AGTGAACGAGATGAGGCTGACCTTGAAGAGTTCGAAATGCGCTTTCGAGAGAGCGGAA 909
QY 784 CAGAGGCCCACTGAGAGGCGCCCTGAAGAGGATTCAGAGTCCGCTTCGAGAGAGCGCGG 843
DB 910 CAGAGGCCCACTGAGAGGCGCCCTGAAGAGGATTCAGAGTCCGCTTCGAGAGAGCGCGG 969
QY 844 GAGGAGGCTCCGCGCAGACTGTGAGGCTCCGAGGCGCAGTGAACCTGTGAGAGGCGCTCTG 903
DB 970 GAGGAGGCTCCGCGCAGACTGTGAGGCTCCGAGGCGCAGTGAACCTGTGAGAGGCGCTCTG 1029
QY 904 CCGATGAAGCAGAGGCCCGGAGAACGATCAGAGGCTGTGTCTCCAGGCTGGGAAAGATAC 963
DB 1030 CCGATGAAGCAGAGGCCCGGAGAACGATCAGAGGCTGTGTCTCCAGGCTGGGAAAGATAC 1089
QY 964 AGAGAGGTAAAGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1023
DB 1090 AGAGAGGTAAAGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1149
QY 1024 -TTGGGCGCAGACCGAGCTGAATGAGAGGTTACGACCCCAAGCCCTGTATTCATTTG 1082
DB 1150 TTGGGCGCAGACCGAGCTGAATGAGAGGTTACGACCCCAAGCCCTGTATTCATTTG 1209
QY 1083 CCGCGAGCTTCACAAACAGGTTTATGATTTGCTGAGCCGCTGTGAAGATGGTTCA 1142
DB 1210 CCGCGAGCTTCACAAACAGGTTTATGATTTGCTGAGCCGCTGTGAAGATGGTTCA 1269
QY 1143 TGGGAGTGTGTGGGCAATTTTGAAGGCTCGAGGAGGCTTTTGAAGAAAGGAGGAGGAG 1202
DB 1270 TGGGAGTGTGTGGGCAATTTTGAAGGCTCGAGGAGGCTTTTGAAGAAAGGAGGAGGAG 1329
QY 1203 CTATATCTGCCCAACTGACCAATTTCTGCAAGTCAAGATGAGATCTTCAGAAACGCG 1262
DB 1330 CTATATCTGCCCAACTGACCAATTTCTGCAAGTCAAGATGAGATCTTCAGAAACGCG 1389
QY 1263 AGATCAGCAGAGAGCTTAATGAGAG---CTTGAGAGTGC-TGATGGCACCGA-TTGTTCA 1317
DB 1390 AGATCAGCAGAGAGCTTAATGAGAGAGCTTGAGAGTGCCTTGATGAGCCGATTTTTC 1449
QY 1318 AGTATGAGAAAT--AGAGCAGAGTCT-----AGCGAAGCAAGGATTAAGGGTGA 1371
DB 1450 AGTATGAGAAATTAAGAGAGAGAGTCTTAAGCGAAGACCCAGAGGTTTAAAGGTGA 1509
QY 1372 ATTGAAGAAAGCTGCAAAATCCAGTGGCAGAGAAACTCAAGATCTTCAGCC----- 1424
DB 1510 ATTGAAGAAAGCTGCAAAATCCAGTGGCAGAGAAACTCAAGATCTTCAGCCCTGATTT 1569
QY 1425 ----- 1424
DB 1570 GGAATTCGTGTCTGAAGACATCTGCTTCACACTCTTGACATGACATTAAGTCA 1629
QY 1425 TGTGATAGAGGCGCTGCTGCTGCTCAAAATGATTTGGCCCGGAGTGTCTCACTGAGCCCA 1484
DB 1630 GGTATATAGAGGCGCTGCTGCTGCTCAAAATGATTTGGCCCGGAGTGTCTCACTGAGCCCA 1689
QY 1485 GCCGAGCTCGGTGTACTGACAGTAAATGACTGTATCTTCAACAGCGCGGAGGAGCAATGAA 1544
DB 1690 GCCGAGCTCGGTGTACTGACAGTAAATGACTGTATCTTCAACAGCGCGGAGGAGCAATGAA 1749

1545 GTTCTTAAGCTCAGGTAAAGAACAGAACCCAAAGCCTTAAGAAAAAGATGAGATGAAGCC 1604
1750 GTTCTTAAGCTCAGGTAAAGAACAGAACCCAAAGCCTTAAGAAAAAGATGAGATGAAGCC 1809
1605 AGAGAACCCAGTCTTCGAAATGCGGTGCTCAGGCGAG 1642
1810 AGAGAACCCAGTCTTCGAAATGCGGTGCTCAGGCTG 1847

RESULT 11
AAZ90579 standard; cDNA; 2867 BP.
AAZ90579;
19-JUN-2000 (first entry)

Murine death inducer-obliterator 1 (DIO-1) polypeptide encoding cDNA.
Death inducer-obliterator 1; DIO-1; cell death; cancer; tumour; murine;
autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
antithrombotic; antiinflammatory; antiproliferative; ss.
Mus sp.
Key Location/Qualifiers
CDS 232..2076
/*tag= a
/product= "DIO-1"

MO200015787-A1.
23-MAR-2000.
10-SEP-1999; 99MO-GB003019.
10-SEP-1998; 98SR-00003069.
17-SEP-1998; 98US-0100873P.
(CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
(BANM/) BANNERMAN D G.
Alonso CM, Domingo DG, Grandlen A, Leonardo B, Martinez P;
MPI; 2000-271426/23.
P-PSDB; AAY67580.
New DNA encoding human and murine death inducer-obliterator 1
polypeptides, useful in the treatment of cancer, autoimmune diseases,
PT diabetes, rheumatoid arthritis, benign tumors, malignant tumors and
hyperproliferative skin disorders.
Claim 3; Fig 1B; 27pp; English.
The invention provides nucleic acids encoding the human and murine death
inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be
expressed by standard recombinant methodology. The DIO-1 polypeptides,
agonists and antagonists are used as a medicament for treating diseases
characterized by an alteration in cell death or by hyperproliferation,
e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
tumours, malignant tumors or hyperproliferative skin disorders. They are
also useful in the treatment of metabolic, proliferative or inflammatory
conditions. The present sequence represents a cDNA encoding the murine
DIO-1 polypeptide
Sequence 2867 BP; 763 A; 666 C; 774 G; 664 T; 0 U; 0 Other;
Query Match 39.2%; Score 1023.6; DB 3; Length 2867;
Best Local Similarity 71.9%; Pred. No. 5.9e-265;
Matches 1494; Conservative 0; Mismatches 534; Indels 50; Gaps 10;
21 TCCGCGCGCTTCGCGAAATGCTCGAGACCTTAAGAGCCTGCGAGACTTACTTCACGCG 80

1 TCCGCTGAGCTCTGGAATGCTGCGGATCCCGCGCGGAGCTTGTTAAGAG 60
81 GAACAGCCTTAGATTAATCTGAGTTGTGAATATCAAGCCTGTACTCTGAACAGTG 140
61 CAGTCCCACTGCTCCCTTGGTTGTTCGAAGCTCGGAATCTTCTATGATGAAGTG 120
141 GCTGACAAACAGTGTGTGTGAGCTGCTGTCTTGTGAACCCAGAGTTTGTCTGCG 200
121 ACTGACAAACAGTGGGTGAG--GCTGCGCTGTCTTGACC-----TGGCCC 166
201 AGGTTTTTGTGTATTTAGGATTT--CAGGAAAAAGTGTCCAGCTTTCAGTGTGAG 259
167 CAGGCTAATTAATTTATGTAGATTTTCAGCCAAAGTTTCCAGCTTTCAGTGTGGA 226
260 CAGGTATGAGACGAACAAAGGACCCGAGCAATGAGAGGACCTAAGGCCATCAACCA 319
227 CAGGTATGAGATTAATTAAGGCACTTGAACATGAGAGACCCAGCTAATCAACCA 286
320 CCAGCAAAAGAGTTCAAGAAAATGAGGTTTTCAGAGACCACTATTCGCAAGCAGAG 379
287 CAGTAAGAGAGTTCAAGAAAATGAGGTTTTCAGAGAACCAAGATTCGCAAGCAGAG 346
380 GCGCAGGGAACCGAGGCTGACCTGAGAGCCGCAACCCCAAGAGAGCTGCGGC 439
347 GTGCAAGAGACAGAGGCGGACCCAGTGAACAGCAAC-----ACAGCAGCATTAAC 400
440 TGTCCCTGCGCGGAGTGTGAGAGGAGCCCAAGCCGCACTGAGGCGGTGAGCAGTTCTGA 499
401 TCTCCCTGCGCGGAGTGTGAGAGGAGCCCAAGCCGCACTGAGGCGGTGAGGTTCTTA 460
500 CCATTGCGCGCGGAGTGTGAGAGGAGCATGCTGTCTCTCTGAGGATTTGAGTAC 559
461 CCAAGGTTCCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 520
560 CCAAGTCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 619
521 CCACATCTTCCACAGTCACTGATGAGACAGCTTCCAGAGGAGGAGGAGGAGGAGGAG 580
620 CTGAGACGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 679
581 CTGAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 637
680 CTGAAAAGGTGAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 739
638 CTGAAAAGGCAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 697
740 TGAAGCTTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 799
698 TTACGTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 757
800 GAGCCCTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 859
758 GGTCCCTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 817
860 AGACTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 919
818 AATCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 877
920 CCGAGAACGATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 979
878 CTGAGGCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 937
980 AGGGAAGGCGGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1039
938 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 997
1040 CTGAATGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1099
998 CTGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1057
1100 AAGGTTTATGATTTTGTCTGTAACCGCTGTGAAGATGTTTCAATGCGATTTGTGGGCA 1159

Db 1058 ACAGTTTATGATCTGCTGATGATGCTGTGAGAGATGGTTCATGCTGACTGTGGTGA 1117
Qy 1160 TTTCTGAGGCTCCAGGAGGCTTTTGGAAAAGATGGGGAAGATATATCTGCCAAACT 1219
Db 1118 TTTCTGAGGCTCCAGGAGGCTCTCTGAAAGAAAGGAGAACTACATCTGCCAAAT 1177
Qy 1220 GCACATTTCTGCAAGTACAGATAGATGATCTCATGAAAACGGCATCAGACAGAGCTA 1279
Db 1178 GCACATTTTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1257
Qy 1280 AATGAGACCTGAGATGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1339
Db 1238 GGTGACATCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1297
Qy 1340 AGTCTAGCAGAAAGCAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTA 1399
Db 1298 AGTCCGAGAAAGCAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTA 1357
Qy 1400 AGAAGAACTCAAGATCTTCCAGCTGTGATGATGATGATGATGATGATGATGATGATG 1459
Db 1358 AGAAGAACTCAAGATCTTCCAGCTGTGATGATGATGATGATGATGATGATGATGATG 1417
Qy 1460 GCCCCGGGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1519
Db 1418 GCCCTGGGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1477
Qy 1520 TCAAAACGCGCGGCAAGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTA 1579
Db 1478 TCAAAACGCGCGGCAAGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTA 1537
Qy 1580 CTAAGAAAGATTAAGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGG 1639
Db 1538 CTAAGAAAGATTAAGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGG 1597
Qy 1640 CAGGATTAAGATTAAGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 1699
Db 1598 TGGGATTAAGATTAAGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 1657
Qy 1700 TGAAGAAAGATTAAGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 1759
Db 1658 TGAAGAAAGATTAAGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 1714
Qy 1760 AGAGCAAGCAGCGCTGCGGCGAGCATCAATTAAGGATTAAGGATTAAGGATTAAGG 1819
Db 1715 AGAGCAAGCAGCGCTGCGGCGAGCATCAATTAAGGATTAAGGATTAAGGATTAAGG 1774
Qy 1820 CTG-----CTGCTCCCTGCGCGCTCACTGTTGATTAAGGATTAAGGATTAAGG 1870
Db 1775 CAGAGAAAGCAGCTGCTGCGCGAGCATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 1834
Qy 1871 TTGGCTCTCTGAGACCTCCCGTCTTCTGATTAAGGATTAAGGATTAAGGATTAAGG 1930
Db 1835 CTGGCTCTCTGAGACCTCCCGTCTTCTGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 1889
Qy 1931 TTGGAGTTGAGAGCTTTGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 1990
Db 1890 CAG-----AGTCTGGGTTGTTCTGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 1940
Qy 1991 CAGAGCCATGTCAGAGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2050
Db 1941 CAGAGCCATGTCAGAGATGCTCTGAGACCCAGAGTGTCTGCTGCTGCTGCTGCTGCTGCT 2000
Qy 2051 TCTGTGATCTGTTTCCAAAGCTGTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2088
Db 2001 CTCTGGGTGTTCTTAAAGAGCTGTGATGAGGCTCATGT 2038

DT 06-NOV-2001 (first entry)
XX
Db Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21863.
XX Human, immune; haematopoietic; immune/haematopoietic antigen; cancer;
Xw Cytosolic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
PD 09-AUG-2001.
PP 17-JAN-2001; 2001WO-US001354.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0230963P.
PR 26-JUL-2000; 2000US-0230964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0237009P.
PR 30-AUG-2000; 2000US-0238924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229533P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.

RESULT 12
ID AAK67051 standard; DNA; 30626 BP.
XX AAK67051;
AC
XX

PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.

PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251909P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Disclosure; SEQ ID NO 21863; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 30626 BP; 7102 A; 7610 C; 7817 G; 8097 T; 0 U; 0 Other;
Query Match 36.3%; Score 948.4; DB 4; Length 30626;
Best Local Similarity 99.6%; Pred. No. 3,7e-244;
Matches 972; Conservative 0; Mismatches 1; Indels 3; Gaps 2;
QY 1636 CAGGCGATTTAAATCTCTTCTGCAAGAGACCGCTCCAAAAAAGAGACC 1695
DB 4785 CAGGCGATTTAAATCTCTTCTGCAAGAGACCGCTCCAAAAAAGAGACC 4844
QY 1696 ACAGTGAAGAGCGAGTGGTGGCTGCGCGAGGAGGACATCGGGAGGAAGCAGCT 1755
DB 4845 ACAGTGAAGAGCGAGTGGTGGCTGCGCGAGGAGGACATCGGGAGGAGGAGCT 4904
QY 1756 TGTGAGAGGAGCAGCGCTGCTGAGGAGGAGGATCAATTAATGATGAGGAGGAG 1815
DB 4905 TGTGAGAGGAGCAGCGCTGCTGAGGAGGAGGATCAATTAATGATGAGGAGGAG 4964
QY 1816 AAGACTGCTGCTCCCTGCGCTGATCTGTTAAATGATGATGATCACTAGGGGTTGGC 1875
DB 4965 AAGACTGCTGCTCCCTGCGCTGATCTGTTAAATGATGATGATCACTAGGGGTTGGC 5024
QY 1876 CTCCTGAGACCCCTCCCTGCTTCTGAGTATGATCCCTGAGGCTGTCAGGACTGGGA 1935
DB 5025 CTCCTGAGACCCCTCCCTGCTTCTGAGTATGATCCCTGAGGCTGTCAGGACTGGGA 5084
QY 1936 GTTGCAAGCTTGTGTTAAGCTATCAAGACACCGGCTCACCATCAAGCGGAGAGAG 1995
DB 5085 GTTGCAAGCTTGTGTTAAGCTATCAAGACACCGGCTCACCATCAAGCGGAGAGAG 5144
QY 1996 CCATGTCAGGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2055
DB 5145 CCATGTCAGGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5204
QY 2056 TCACGTTTTCAGGAGCTGTAACCTCACTGGGAGAGGTTTCACTTAATGATGATTTCT 2115
DB 5205 TCACGTTTTCAGGAGCTGTAACCTCACTGGGAGAGGTTTCACTTAATGATGATTTCT 5264

QY 2116 TAATCTGTTTCACTCTGAGGCTCTGTAAGATTTGATTTCTTCTCATCCAGTCTG 2175
DB 5665 TAACTCTGTTTCACTCTGAGGCTCTGTAAGATTTGATTTCTTCTCATCCAGTCTG 5324
QY 2176 ATTGCATAGCCACACTGCCCCGCGACGCCACATCCACCCCTGCTGACATGAGTTGTTCT 2235
DB 5325 ATTGCATAGCCACACTGCCCCGCGACGCCACATCCACCCCTGCTGACATGAGTTGTTCT 5384
QY 2236 GACAAACGCGCTGATAGCTTCACTTTTTCACATTTCTCCACGCGCGACACATGAAAGC 2295
DB 5385 GACAAACGCGCTGATAGCTTCACTTTTTCACATTTCTCCACGCGCGACACATGAAAGC 5444
QY 2296 ATCACTCTTTTATGTTGAGGAACTTTGCAAGTTGATGTTGATGTTTTCAGG 2355
DB 5445 ATCACTCTTTTATGTTGAGGAACTTTGCAAGTTGATGTTGATGTTTTCAGG 5504
QY 2356 TGTACATTTATTTTGAATGCTGCGGAGATGAGGAAATTTTTTTTTCATGTCGATTC 2415
DB 5505 TGTACATTTATTTTGAATGCTGCGGAGATGAGGAAATTTTTTTTTCATGTCGATTC 5562
QY 2416 CACCTTACACACCCACATGAAACATTTGAACTTGAAGGCGACACTCTGCTTCACTA 2474
DB 5563 CACCTTACACACCCACATGAAACATTTGAACTTGAAGGCGACACTCTGCTTCACTA 5622
QY 2475 GAGCCCAACGTAAGTGAAGTCAACCTAGAACACTGCTCTGACCGACGAGCGCTGCTT 2534
DB 5623 GAGCCCAACGTAAGTGAAGTCAACCTAGAACACTGCTCTGACCGACGAGCGCTGCTT 5682
QY 2535 GAGCTTGTATTTCTACATGATGATCTGCTTCTGCTGCTCTGTAATGTTAGACTC 2594
DB 5683 GAGCTTGTATTTCTACATGATGATCTGCTTCTGCTGCTCTGTAATGTTAGACTC 5742
QY 2595 TTAGATCATATCTCTG 2610
DB 5743 TTAGATCATATCTCTG 5758

RESULT 13

AAS86575 standard; cDNA. 2986 BP.

AAS86575;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #22379.

Human; chromosome mapping; gene mapping; gene therapy; forensic;

food supplement; medical imaging; diagnostic; genetic disorder; ss.

Homo sapiens.

MO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001MO-US008631.

31-MAR-2000; 2000US-00540217.

23-AUG-2000; 2000US-00649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

P-PSDB; ABG22388.

New isolated polynucleotide and encoded polypeptides, useful in

diagnostics, forensics, gene mapping, identification of mutations

responsible for genetic disorders or other traits and to assess

biodiversity.

PS Claim 1; SEQ ID NO 22379; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (II) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful for medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences
XX

Sequence 2986 BP; 798 A; 796 C; 781 G; 611 T; 0 U; 0 Other;

Query Match 26.2%; Score 682.6; DB 5; Length 2986;

Best Local Similarity 88.1%; Pred. No. 5.3e-173; Mismatches 4; Indels 116; Gaps 9;

Matches 891; Conservative 0; Mismatches 4; Indels 116; Gaps 9;

965 GAGAGATGATTTGAGGAAAGGCGCTCAGACATCAAGATGAGAGCTGAGACT 1024

1 GAGAGATGATTTGAGGAAAGGCGCTCAGACATCAAGATGAGAGCTGAGACT 60

1025 TGGGCGACCCGAGGCTGAAATGTAAGGTTAGACCCGAGCGCTGATGAGATGTTGCT 1084

61 TGGGCGACCCGAGGCTGAAATGTAAGGTTAGACCCGAGCGCTGATGAGATGTTGCT 120

1085 GCGAGCTCACAACAAGGTTTATGATTTGATGACCGCTGTAAGATGTTGCT 1144

121 GCGAGCTCACAACAAGGTTTATGATTTGATGACCGCTGTAAGATGTTGCT 180

1145 GCGATTTGTTGGGCAATTTCTAGAGCTGAGGAGGCTTTTGAAGAAAGATGTTGCT 1204

181 GCGATTTGTTGGGCAATTTCTAGAGCTGAGGAGGCTTTTGAAGAAAGATGTTGCT 240

1205 ATATTTGCCAAATGTCACCAATTTCTGCAAGTGCAGAGATGAGACTATTCAGAAAGCGCG 1264

241 ATATTTGCCAAATGTCACCAATTTCTGCAAGTGCAGAGATGAGACTATTCAGAAAGCGCG 300

1265 ATGAGCAGAAAGCTAAATGAGAGCTGAGATGCTGATGCGACCGATTTGTAAGATGAG 1324

301 ATGAGCAGAAAGCTAAATGAGAGCTGAGATGCTGATGCGACCGATTTGTAAGATGAG 360

1325 GAAATATGAGCAGAGAGTCTAGCGAAGCAAGGATTAAGGTTGTAAGAAAGCTG 1384

361 GAAATATGAGCAGAGAGTCTAGCGAAGCAAGGATTAAGGTTGTAAGAAAGCTG 420

1385 CAATTCAGTGGCAGAAAGAACTCAGATCTTCCAGCT 1425

421 CAATTCAGTGGCAGAAAGAACTCAGATCTTCCAGCT 480

1426 1425

481 CCCAGCTGCTGTCTCTGCGAGGATTTGGAATTTCTGTGTTGAAAGCATCTGCT 540

1426 GTGATAGAGCGCTGCTGCTCAAAATGTA 1456

541 TCACACTTTGACATGATTAAGCTCAAGGTAATGAGCGCTGTCTCAAAATGTA 600

1457 TTGGCCCCCGGCTGCTGACGTTGCGAGCGCCGAGCTGCTGCTCAAAATGTA 1513

601 TTGGCCCCCGGCTGCTGACGTTGCGAGCGCCGAGCTGCTGCTCAAAATGTA 660

XX 07-FEB-2001.
 PD 28-JUL-2000; 2000EP-00116126.
 XX 28-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX (HELI-) HELIX RES INST.
 PA Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 DR Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.
 XX
 PS Claim 1; SEQ ID NO 4840; 2537bp + Sequence Listing: English.
 CC The present invention describes primer sets for synthesizing 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH01166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX
 SO Sequence 681 BP; 161 A; 177 C; 220 G; 119 T; 0 U; 4 Other;
 Query Match 19.6%; Score 510.8; DB 4; Length 681;
 Best Local Similarity 98.9%; Pred. No. 5.3e-127;
 Matches 534; Conservative 0; Mismatches 4; Indels 2; Gaps 2;
 QY 65 GAGGTTTCTCCACGCGAAACGCTCTAGATATCTGAGTTTGTGAAAATACGAAGCTG 124
 DB 101 GAGGTTTCTCCACGCGAAACGCTCTAGATATCTGAGTTTGTGAAAATACGAAGCTG 160
 QY 125 TTAATCGTGAACAGTGGCTGACAAACAGTGTGTTGAGCCCTGCTGCTGAGACCC 184
 DB 161 TTAATCGTGAACAGTGGCTGACAAACAGTGTGTTGAGCCCTGCTGCTGAGACCC 220
 QY 185 AGAGGTTTCTGCTGCAGGCTTTTGTGTTGATTTAGATTTCAAGGAAAAGTGTCCAG 244
 DB 221 AGAGGTTTCTGCTGCAGGCTTTTGTGTTGATTTAGATTTCAAGGAAAAGTGTCCAG 280
 QY 245 CTTTACAGTTCGACAGTATGACGACAAAGGCGACCCGAGCAATGAGAGGACCTA 304
 DB 281 CTTTACAGTTCGACAGTATGACGACAAAGGCGACCCGAGCAATGAGAGGACCTA 340
 QY 305 AGCCATCAAAACCCAC-AGCAAGAGTTCAGAAAAACATGGGGTTCGAGAGGACCT 363
 DB 341 AGCCATCAAAACCCACAGCAAGAGTTCAGAAAAACATGGGGTTCGAGAGGACCT 400

QY 364 ATGSCCAAGGAGAGGGGCGAGGGGAGCGGAGGCTGACCCACTGAGCCGCCACCCCA 423
 DB 401 ATGCCCAAGGAGAGGGGCGAGGGGAGCGGAGGCTGACCCACTGAGCCGCCACCCCA 460
 QY 424 CAGCAGAGCTGGGCTGTCTCTGCGGCGAGTGGAGGCGAGCCCAAGCGCACTGAGCGC 483
 DB 461 CAGCAGAGCTGGGCTGTCTCTGCGGCGAGTGGAGGCGAGCCCAAGCGCACTGAGCGC 520
 QY 484 GTGAGAGATTCTGACATTTGCGGCGCGCGCGGAGAGAGCATGCTGCTCCCTG 543
 DB 521 GTGAGAGATTCTGACATTTGCGGCGCGCGGAGAGAGCATGCTGCTCCCTG 580
 QY 544 GAGGATCT-GGTGGGCCAGGTCCTGCGCCGACAGACCGGAGACAGGCTCCGAGG 602
 DB 581 GAGGATCTGGGTTAGCCCACTCTGCGCCGACAGACCGGAGACAGGCTCCGAGG 640

Search completed: April 23, 2004, 08:28:12
 Job time : 983.654 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 04:42:12 ; Search time 9937.73 Seconds
(without alignments)
11383.418 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 2610
Sequence: 1 ctgcgtgcgcgcgcgcgcac.....actcttaagatcatatccctg 2610

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBml:*

1: gb_ba:*

2: gb_hcg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pac:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_rfc8:*

12: gb_ey:*

13: gb_un:*

14: gb_vt:*

15: em_da:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pac:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_rfc8:*

28: em_un:*

29: em_vt:*

30: em_hcg_hum:*

31: em_hcg_in:*

32: em_hcg_om:*

33: em_hcg_ov:*

34: em_hcg_pac:*

35: em_hcg_ph:*

36: em_hcg_pl:*

37: em_hcg_pr:*

38: em_hcg_ro:*

39: em_hcg_rfc8:*

40: em_hcg_un:*

41: em_hcg_vt:*

42: em_hcg_om_hum:*

43: em_hcg_om_in:*

44: em_hcg_om_mu:*

45: em_hcg_om_om:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	2610	100.0	2610	6	BD244712
2	2610	100.0	2610	6	AX023369
3	2532.4	97.0	2725	9	BC014489
4	2517.8	96.5	2767	9	BC004237
5	2516.2	96.4	2772	6	AX878224
6	2516.2	96.4	2772	6	BD157094
7	2516.2	96.4	2772	9	AK002127
8	1941	74.4	2407	9	BC000770
9	1023.6	39.2	2602	10	MM0238332
10	1023.6	39.2	2602	10	BD244713
11	1023.6	39.2	2602	6	AX023370
12	985	37.7	4906	10	AK129117
13	948.4	36.3	160241	9	HS885L7
14	770.6	29.5	6692	9	AB002331
15	654.6	25.1	28519	9	HS8563B14
16	596.8	22.9	659	6	AX897185
17	596.8	22.9	659	6	BD032718
18	510.8	19.6	681	6	AX869935
19	510.8	19.6	681	6	BD149997
20	491.4	18.8	650	9	HMMVPI0D03
21	490.4	18.8	272404	2	AC108337
22	481.8	18.5	104653	10	AL732560
23	449.8	17.2	221547	2	AC119716
24	386.4	14.8	69252	2	AC101519
25	336.6	12.9	542	6	AX874900
26	336.6	12.9	542	6	BD154962
27	262.4	10.1	2428	5	BC060442
28	262.4	8.9	69252	2	AC101519
29	202.8	7.8	472	6	AR426988
30	202.8	7.8	472	6	BD122541
31	103.4	4.0	198133	3	AK114522
32	70.8	2.7	2161	2	AC017132
33	68.6	2.6	61204	2	AC007594
34	68.6	2.6	161601	3	AC007594
35	68.6	2.6	225655	3	AB003695
36	67	2.6	7218	6	166494
37	66.2	2.5	125020	9	AF429315
38	65	2.5	2888	9	AK055678
39	65	2.5	5975	6	AX329603
40	65	2.5	6256	6	AR338778
41	65	2.5	6936	9	D87685
42	65	2.5	6948	6	AX210673
43	65	2.5	6948	9	AF091622
44	65	2.5	7142	9	HS8608416
45	64	2.5	125020	9	AF429315

ALIGNMENTS

RESULT 1
BD244712
LOCUS
DEFINITION
Gene encoding for the human and murine death
inducer-obliator-1.
ACCESSION
BD244712
VERSION
BD244712.1
KEYWORDS
JP 2002526040-A/1.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
Alonso, C.M., Domingo, D.G., Grandien, A., Leonardo, E. and Martinez, P.
1 (bases 1 to 2610)
TITLE
Genes encoding for the human and murine death inducer-obliator-1

JOURNAL
Patent: JP 2002526040-A 1 20-AUG-2002;
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS
OS Homo sapiens (human)
COMMENT
JP 2002526040-A/1
PD 20-AUG-2002
PR 10-SEP-1999 JP 2000570314
PR 10-SEP-1998 SE 9803069-5,17-SEP-1998 US 60/100873 P1
CARLOS MARTINEZ ALONSO, DAVID GARCIA DOMINGO, ALF GRANDIEN, P1
ESTHER LEONARDO,
P1 PEDRO MARTINEZ
PC C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P3/10,A61P17/00, PC
A61P29/00,
PC A61P35/00,A61P37/06,A61P43/00,C07K14/47,C07K16/18,C12N1/21, PC
C12N5/10,
PC C12P21/02,C12Q1/02,C12Q1/68,G01N33/15,G01N33/50//C12P21/08, PC
C12N15/00,
PC C12N5/00,A61K37/02
CC This gene is referred to in the application as Human Death
Inducer
CC Obliterator Gene 1.
CC It has been named by the Human Gene
Nomenclature Committee
CC as
CC Death-Associated Transcription Factor (DAF-1) FH Key
CC Location/Qualifiers
FT source 1..2610
FT Location/Qualifiers
source 1..2610
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 2610; DB 6; Length 2610;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2610; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTGGTGGCCGTCGCCCACTCCGCGGCTTCGGGGAAATGGCTGCCAGACCTTGAAGGC 60
DB 1 CTGGTGGCCGTCGCCCACTCCGCGGCTTCGGGGAAATGGCTGCCAGACCTTGAAGGC 60
QY 61 CTGCGAGCTTACTCCACGGGAAAGAGCTCTAGATAATCTGAGTTGTTGAAATATCAAG 120
DB 61 CTGCGAGCTTACTCCACGGGAAAGAGCTCTAGATAATCTGAGTTGTTGAAATATCAAG 120
QY 121 CTGTACTCGTGAACAGTGGCTGACAAAGTGTGTGTGAGGCTGCTGTCTGCTTGG 180
DB 121 CTGTACTCGTGAACAGTGGCTGACAAAGTGTGTGTGAGGCTGCTGTCTGCTTGG 180
QY 181 ACCGAGAGTTTGGCTCTGCCAGGGTTTTGTTGTATTTAGGATTTCAAGGAAAGTGTCT 240
DB 181 ACCGAGAGTTTGGCTCTGCCAGGGTTTTGTTGTATTTAGGATTTCAAGGAAAGTGTCT 240
QY 241 CAAGCTTCACTGTGTGAGCAGGTATGAGCAAAAGGCGACCGAGCAATGAGAGGCA 300
DB 241 CAAGCTTCACTGTGTGAGCAGGTATGAGCAAAAGGCGACCGAGCAATGAGAGGCA 300
QY 301 CCTAAGGCAATCAAAACCAACAGCAAAAGTTCAGAAAAATGAGGTTTTGAAAGACC 360
DB 301 CCTAAGGCAATCAAAACCAACAGCAAAAGTTCAGAAAAATGAGGTTTTGAAAGACC 360
QY 361 ACTATGCCCAAGAGAGAGGGGCGCAGGGGAGCTGACCACTGAGAGCCCAAGCC 420
DB 361 ACTATGCCCAAGAGAGAGGGGCGCAGGGGAGCTGACCACTGAGAGCCCAAGCC 420
QY 421 CCAAGCAGAGCAGTGGCTGTCTGCGGCGCAGTGGAGAGCAGCCCAAGCAGCACTGAG 480
DB 421 CCAAGCAGAGCAGTGGCTGTCTGCGGCGCAGTGGAGAGCAGCCCAAGCAGCACTGAG 480
QY 481 CGGCTGAGAGATTTCTGACCAATTTGCGCGGCGCGGCGAGAGAGACATGCTGTCTCC 540
DB 481 CGGCTGAGAGATTTCTGACCAATTTGCGCGGCGCGGCGAGAGAGACATGCTGTCTCC 540

QY 541 CTGAGAGATTTCTGAGAGCCCAAGCTTCGCGGCGCAAGACGCGAGACAGCTTCGAG 600
DB 541 CTGAGAGATTTCTGAGAGCCCAAGCTTCGCGGCGCAAGACGCGAGACAGCTTCGAG 600
QY 601 GCGAGCGTGAAGACGCTTCTGAGACCAAGAGCGGCCCCAGTCTGCTTCCACAGCTGTG 660
DB 601 GCGAGCGTGAAGACGCTTCTGAGACCAAGAGCGGCCCCAGTCTGCTTCCACAGCTGTG 660
QY 661 AAGGAGAGCAAGCTCTTCTGAAAAGGTGAAGAAGGGGATGACAGATGACCTTCC 720
DB 661 AAGGAGAGCAAGCTCTTCTGAAAAGGTGAAGAAGGGGATGACAGATGACCTTCC 720
QY 721 GATAGTGAACAGCGATGGCTGACCTTTGAAAAGAGCTTCAGAAATGCGCTTCGAGAGAGCG 780
DB 721 GATAGTGAACAGCGATGGCTGACCTTTGAAAAGAGCTTCAGAAATGCGCTTCGAGAGAGCG 780
QY 781 GAACAGAGGCCCACTGAGAGGCCCCGAGAAAGGATCCAGAGTCCGCTTCGAGAGAGAGCG 840
DB 781 GAACAGAGGCCCACTGAGAGGCCCCGAGAAAGGATCCAGAGTCCGCTTCGAGAGAGAGCG 840
QY 841 CGGAGAGAGGCTCCCGCGAGACTGTGGGCTCCGAGGCGCACTGTGAGAGAGGCGCTC 900
DB 841 CGGAGAGAGGCTCCCGCGAGACTGTGGGCTCCGAGGCGCACTGTGAGAGAGGCGCTC 900
QY 901 CTGCGCAGTAAAGCAGAGGCCCCGAGAAAGATCAAGGGGGTGTGTCGAGGCTGGGAAAGAT 960
DB 901 CTGCGCAGTAAAGCAGAGGCCCCGAGAAAGATCAAGGGGGTGTGTCGAGGCTGGGAAAGAT 960
QY 961 GACAGAGAGTAAAGTTGAGAGGAAAGCGGCTCAGAGCATCAAAAGATGAGAGGCTGGA 1020
DB 961 GACAGAGAGTAAAGTTGAGAGGAAAGCGGCTCAGAGCATCAAAAGATGAGAGGCTGGA 1020
QY 1021 GACTTGGGCGGAGCCGAAGCTCGAATGTGAGGGTTACGACCCCAAGCCCTGTATTGCAATT 1080
DB 1021 GACTTGGGCGGAGCCGAAGCTCGAATGTGAGGGTTACGACCCCAAGCCCTGTATTGCAATT 1080
QY 1081 TGCCGCGAGGCTCAACAACAGGTTTATGATTGTGCTGTGACCGCTGTGAAGATGTTTT 1140
DB 1081 TGCCGCGAGGCTCAACAACAGGTTTATGATTGTGCTGTGACCGCTGTGAAGATGTTTT 1140
QY 1141 CATGGCAGTTGTGTGGCAATTTCTGAGGCTCGAGGAGGCTTTTGAAGGAATGGGAA 1200
DB 1141 CATGGCAGTTGTGTGGCAATTTCTGAGGCTCGAGGAGGCTTTTGAAGGAATGGGAA 1200
QY 1201 GACTATATCTGCCAACTGCAATTTCTGCAAGTGAAGATGAGACTCATTCAGAAACG 1260
DB 1201 GACTATATCTGCCAACTGCAATTTCTGCAAGTGAAGATGAGACTCATTCAGAAACG 1260
QY 1261 GCAGATAGCAGAGAGCTTAAATGAGACCTGAGAGATGCTGATGAGCAATTTGACAGT 1320
DB 1261 GCAGATAGCAGAGAGCTTAAATGAGACCTGAGAGATGCTGATGAGCAATTTGACAGT 1320
QY 1321 ATAGGAACAATAGAGCAGAGTCTAGGAGAGCCAGAGGATTAAGGTTTGAAGAA 1380
DB 1321 ATAGGAACAATAGAGCAGAGTCTAGGAGAGCCAGAGGATTAAGGTTTGAAGAA 1380
QY 1381 GCTGCAAAATCCAGTGGCAAGAGAACTCAAGATCTTCCAGCTGTGATGAGGCGCTT 1440
DB 1381 GCTGCAAAATCCAGTGGCAAGAGAACTCAAGATCTTCCAGCTGTGATGAGGCGCTT 1440
QY 1441 GGTGCGCTCAAAATGATTTGGCCCGGGTGTGTGACGTGGGGCGAGCCGATCGGTATC 1500
DB 1441 GGTGCGCTCAAAATGATTTGGCCCGGGTGTGTGACGTGGGGCGAGCCGATCGGTATC 1500
QY 1501 TGCAATATGACATGATCTCTCAAAACAGCGGAGGCAATGAGATTTCTAAGCTCAGGT 1560
DB 1501 TGCAATATGACATGATCTCTCAAAACAGCGGAGGCAATGAGATTTCTAAGCTCAGGT 1560
QY 1561 AAGAGACAGAAAGCCAAAGCTTAAAGAAAGATGAAGATGAAGCCAGAGAGCCAGTCTT 1620
DB 1561 AAGAGACAGAAAGCCAAAGCTTAAAGAAAGATGAAGATGAAGCCAGAGAGCCAGTCTT 1620

Qy	1621	CCGAATATCGGTGCTCAGCCAGGATTTAAATCTCTTCTGTGCACAAGACCCAGCTCCA	1680
Db	1621	CCGAATATCGGTGCTCAGCCAGGATTTAAATCTCTTCTGTGCACAAGACCCAGCTCCA	1680
Qy	1681	GAATAAANAAGACCAACATGTAAGAAGGACAGTGTGCTCCCTGCGGAGTGAAGACATC	1740
Db	1681	GAATAAANAAGACCAACATGTAAGAAGGACAGTGTGCTCCCTGCGGAGTGAAGACATC	1740
Qy	1741	GGGAAGGAGACAGCTTGTGAGAGACAGCCGCTGTGTGAGGAGCATCACAAATTACAAT	1800
Db	1741	GGGAAGGAGACAGCTTGTGAGAGACAGCCGCTGTGTGAGGAGCATCACAAATTACAAT	1800
Qy	1801	GCAGTAAAGCCAGAAAAGAATGTCGTCCCTGCGGCTCACTGTGTATTAATGTATGAT	1860
Db	1801	GCAGTAAAGCCAGAAAAGAATGTCGTCCCTGCGGCTCACTGTGTATTAATGTATGAT	1860
Qy	1861	CACCTAAGGGGTGGCTCTCTGGAACCCCTCCGTTCTTCTGGAATAGCATCCCTGAGCC	1920
Db	1861	CACCTAAGGGGTGGCTCTCTGGAACCCCTCCGTTCTTCTGGAATAGCATCCCTGAGCC	1920
Qy	1921	TGTCAGAGACTGGAGTTGCCAGCTTTGTGTTAAGCTGATCAACAACCCGCTGACAT	1980
Db	1921	TGTCAGAGACTGGAGTTGCCAGCTTTGTGTTAAGCTGATCAACAACCCGCTGACAT	1980
Qy	1981	CAGCGGGGAAGGAGAGCCCATGTCAGAGATGCTCTGTCGCGGCTGATCCCTAGTGC	2040
Db	1981	CAGCGGGGAAGGAGAGCCCATGTCAGAGATGCTCTGTCGCGGCTGATCCCTAGTGC	2040
Qy	2041	TGTCAGAGACTTCTGTCACTGTGTTTCCAAAGCTGTAAACCTCACTGTGTGAAGTTCACT	2100
Db	2041	TGTCAGAGACTTCTGTCACTGTGTTTCCAAAGCTGTAAACCTCACTGTGTGAAGTTCACT	2100
Qy	2101	TAAATGATTGATTTCTTTAAATCTCTGTTTCACTCAAGCTCTGTGATGATTTGATTC	2160
Db	2101	TAAATGATTGATTTCTTTAAATCTCTGTTTCACTCAAGCTCTGTGATGATTTGATTC	2160
Qy	2161	CTTCATCCCAAGTCTGATGTGCAATAGCCACACTGCCCGGACCCCACTCACCCCTGTCTG	2220
Db	2161	CTTCATCCCAAGTCTGATGTGCAATAGCCACACTGCCCGGACCCCACTCACCCCTGTCTG	2220
Qy	2221	CACATGAGTTGTGTGACAAACAGGCGCTGTATAGCCTTCAGTTTTTCCACATGTGCCAGG	2280
Db	2221	CACATGAGTTGTGTGACAAACAGGCGCTGTATAGCCTTCAGTTTTTCCACATGTGCCAGG	2280
Qy	2281	CCAGACATGAAGAACATCACTTCTTTTAAAGTGTGGGAATCTTGCAAGTAAAGTGTG	2340
Db	2281	CCAGACATGAAGAACATCACTTCTTTTAAAGTGTGGGAATCTTGCAAGTAAAGTGTG	2340
Qy	2341	CATCTGATTTTCAAGTGTACATTTATTTTGAATGAGGGAATCTTGCAAGTAAAGTGTG	2400
Db	2341	CATCTGATTTTCAAGTGTACATTTATTTTGAATGAGGGAATCTTGCAAGTAAAGTGTG	2400
Qy	2401	TCCATGTCCGATTCACACGCTACACACCCACATGAACATTTGCAATCTTGCAAGGACATC	2460
Db	2401	TCCATGTCCGATTCACACGCTACACACCCACATGAACATTTGCAATCTTGCAAGGACATC	2460
Qy	2461	ACTCTGTCTTATAGGCCCCCAAGGTAAAGTGAATTCACCTAAGAACATGTCTCTGACCGC	2520
Db	2461	ACTCTGTCTTATAGGCCCCCAAGGTAAAGTGAATTCACCTAAGAACATGTCTCTGACCGC	2520
Qy	2521	AGGAGCGGTGCTTGAATTTGATTTTCAATGATGATCTGCTTCTTGCCCTGTCTCTT	2580
Db	2521	AGGAGCGGTGCTTGAATTTGATTTTCAATGATGATCTGCTTCTTGCCCTGTCTCTT	2580
Qy	2581	GAATGTTTAACTCTTAAAGATCAATATCTG 2610	
Db	2581	GAATGTTTAACTCTTAAAGATCAATATCTG 2610	

RESULT 2			
AX023369			
LOCUS	AX023369	2610 bp	DNA
DEFINITION	Sequence 1 from Patent W00015787.	linear	PAT 15-SBP-2000

Accession	Version	Keywords	Source	Organism	Reference Authors Title	Journal	Features	Source
AX023369	AX023369.1	GI:10183781		Homo sapiens (human)	Leonardo, B., Martinez, P., Alonso, C.M., Domingo, D.G. and Grandien, A. Genes encoding for the human and murine death in ducer-oblietator-1	Patent: WO 0015787-A 1 23-MAR-2000;		
				Homo sapiens	Bukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	LEONARDO ESTHER (ES) ; DOMINGO DAVID GARCIA (DS) ; MARTINEZ ALONSO CARLOS (ES) ; BANNERMAN DAVID GARDNER (GB) ; GRANDIEN ALF (SE)		
						Location/Qualifiers		
						1. 2610		
						/organism="Homo sapiens"		
						/mol_type="unassigned DNA"		
						/db_xref="taxon:9606"		
						/note="This gene is referred to in the application as Human Death Inducer Obliterator Gene 1. It has now been named by the Human Gene Nomenclature Committee as Death-Associated Transcription Factor (DATF-1)"		
Query Match	100.0%	Score 2610;	DB 6;	Length 2610;				
Best Local Similarity	100.0%	Pred. No. 0;						
Matches 2610;	Conservative 0;	Mismatches	0;	Indels	0;	Gaps	0;	
QY	1	CTGGTGGCCGCTCCGACCTCCGCGCGCTTGGGGAATAGGCGGAGACCTTAGAGGC	60					
DB	1	CTGGTGGCCGCTCCGACCTCCGCGCGCTTGGGGAATAGGCGGAGACCTTAGAGGC	60					
QY	61	CTGGGAGACTTACTCCACGGGAAACAGGCTTAGATATCTGAGTTGTTGAAATACGAG	120					
DB	61	CTGGGAGACTTACTCCACGGGAAACAGGCTTAGATATCTGAGTTGTTGAAATACGAG	120					
QY	121	CTGTGTAATCGTGAACAGTGGCTGACCAACAGTGTGTTGTAAGCTCGGCTGTCTGTTGG	180					
DB	121	CTGTGTAATCGTGAACAGTGGCTGACCAACAGTGTGTTGTAAGCTCGGCTGTCTGTTGG	180					
QY	181	ACCCAGAGGTTTGTCTGCGCAGGGTTTTGGTGTATTTAGATTTCAGGGAATAGTCTC	240					
DB	181	ACCCAGAGGTTTGTCTGCGCAGGGTTTTGGTGTATTTAGATTTCAGGGAATAGTCTC	240					
QY	241	CAAGCTTCAAGTGTGGAGAGGATATGAGACAAAGAGCGACCCGACCAATGAGAGGCA	300					
DB	241	CAAGCTTCAAGTGTGGAGAGGATATGAGACAAAGAGCGACCCGACCAATGAGAGGCA	300					
QY	301	CCTAAGGCATCAAAACCCACACAGCAAAAGATTACGAGAAAATAGGGGTTTTGAGGACC	360					
DB	301	CCTAAGGCATCAAAACCCACACAGCAAAAGATTACGAGAAAATAGGGGTTTTGAGGACC	360					
QY	361	ACTATGCGCAAGGAGAGGCGCGAGGAGATCGCGAGGCTTACCCATCTGAGGCGCCACC	420					
DB	361	ACTATGCGCAAGGAGAGGCGCGAGGAGATCGCGAGGCTTACCCATCTGAGGCGCCACC	420					
QY	421	CCACAGAGAGAGTGGGCTGTCCCTGCGCGCGAGTGGAGGGACCCCAAGCCGACCTGAG	480					
DB	421	CCACAGAGAGAGTGGGCTGTCCCTGCGCGCGAGTGGAGGGACCCCAAGCCGACCTGAG	480					
QY	481	CGCTGAGAGAGTTCTGAGCAATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	540					
DB	481	CGCTGAGAGAGTTCTGAGCAATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	540					
QY	541	CTGAGAGATTCGCTGAGGCCCAAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	600					
DB	541	CTGAGAGATTCGCTGAGGCCCAAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	600					
QY	601	GGAGAGCTGGAAGAGCGCTTCTGAGACCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCG	660					
DB	601	GGAGAGCTGGAAGAGCGCTTCTGAGACCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCG	660					

```

Qy 661 AAGGAACGACGAGCTCTTCTGAAAGGTGAAGAGGGGATGACCAAGATGACACCTCC 720
Db 661 AAGGAACGACGAGCTCTTCTGAAAGGTGAAGAGGGGATGACCAAGATGACACCTCC 720
Qy 721 GATAGTACAGCAGATGAGCTGACCTTGAAGAGCTTCAAGATGAGCTTTCAGAGAGGGG 780
Db 721 GATAGTACAGCAGATGAGCTGACCTTGAAGAGCTTCAAGATGAGCTTTCAGAGAGGGG 780
Qy 781 GAAACGAGACCCATCTGAGAGGCCCCCTGAAAGGGATCCAGAGTCCCTGCGGAAAGAGCC 840
Db 781 GAAACGAGACCCATCTGAGAGGCCCCCTGAAAGGGATCCAGAGTCCCTGCGGAAAGAGCC 840
Qy 841 CGGAGAGAGGGTCCCGCGAGACTGTGGGCTCCAGAGCCAGTGAACATGTGAGAGGGCTC 900
Db 841 CGGAGAGAGGGTCCCGCGAGACTGTGGGCTCCAGAGCCAGTGAACATGTGAGAGGGCTC 900
Qy 901 CTGCCAGTAAAGCAGAGAGCCCGAGAGACGATCAGAGGGGTGTGTCTCCAGAGCTGGGAAAGAT 960
Db 901 CTGCCAGTAAAGCAGAGAGCCCGAGAGACGATCAGAGGGGTGTGTCTCCAGAGCTGGGAAAGAT 960
Qy 961 GACAGAGAGTAAAGTGAAGAGGAAAGGCGCTCAGACATCAAGATGAGAGAGCTTGA 1020
Db 961 GACAGAGAGTAAAGTGAAGAGGAAAGGCGCTCAGACATCAAGATGAGAGAGCTTGA 1020
Qy 1021 GACTTGGGCGGACGGAAGCCGGAAGTGAAGGGTTACGACCCCAAGCCCTGTATTTGCA 1080
Db 1021 GACTTGGGCGGACGGAAGCCGGAAGTGAAGGGTTACGACCCCAAGCCCTGTATTTGCA 1080
Qy 1081 TGCCGCGAGCTCTCAACAACAAGGTTTATGATTTGCTGTGACCGCTGTGAAGATGAT 1140
Db 1081 TGCCGCGAGCTCTCAACAACAAGGTTTATGATTTGCTGTGACCGCTGTGAAGATGAT 1140
Qy 1141 CATGGCGATTTGTGGGCAATTTCTGAGGCTCGAGGAGGCTTTTGAAGAGATGGGGA 1200
Db 1141 CATGGCGATTTGTGGGCAATTTCTGAGGCTCGAGGAGGCTTTTGAAGAGATGGGGA 1200
Qy 1201 GACTATATCTGCGCAAACTGCAACATCTGCAAGTGCAGAGTGAAGTCAATTCAGAAAG 1260
Db 1201 GACTATATCTGCGCAAACTGCAACATCTGCAAGTGCAGAGTGAAGTCAATTCAGAAAG 1260
Qy 1261 GCAGATCAGCAGAGAGCTTAAATGAGAGCTGAGAGTGTGATGAGCAGATTTGACAG 1320
Db 1261 GCAGATCAGCAGAGAGCTTAAATGAGAGCTGAGAGTGTGATGAGCAGATTTGACAG 1320
Qy 1321 ATAGGAACAATGAGCAGAGAGCTTAAATGAGAGCTGAGAGTGTGATGAGCAGATTTG 1380
Db 1321 ATAGGAACAATGAGCAGAGAGCTTAAATGAGAGCTGAGAGTGTGATGAGCAGATTTG 1380
Qy 1381 GCTGCAATTCAGAGTGCAGAGAGAACTCAAGATCTTCAAGCTGTGATGAGAGCGCT 1440
Db 1381 GCTGCAATTCAGAGTGCAGAGAGAACTCAAGATCTTCAAGCTGTGATGAGAGCGCT 1440
Qy 1441 GGTGCTCTCAAAATGATTTGCCCCCGGTGTCTGTCACTGTGCGCAGCCGACTCGTGTAC 1500
Db 1441 GGTGCTCTCAAAATGATTTGCCCCCGGTGTCTGTCACTGTGCGCAGCCGACTCGTGTAC 1500
Qy 1501 TGCAGTAAATGATCTGTATCTTCAACACGCGCGAGAGATGAAATTTCTAAGCTCAGGT 1560
Db 1501 TGCAGTAAATGATCTGTATCTTCAACACGCGCGAGAGATGAAATTTCTAAGCTCAGGT 1560
Qy 1561 AAAGGAACGAGAGCCAAAGCCTTAAAGAAAGATGAAGAGCCAGAGAGCCAGTCTT 1620
Db 1561 AAAGGAACGAGAGCCAAAGCCTTAAAGAAAGATGAAGAGCCAGAGAGCCAGTCTT 1620
Qy 1621 CCGAATATCGGTGCTCAGGCAAGTATTAATCTTCTGTGCAACAAGAGCAGCTCCA 1680
Db 1621 CCGAATATCGGTGCTCAGGCAAGTATTAATCTTCTGTGCAACAAGAGCAGCTCCA 1680
Qy 1681 GAAAAAAGAGAGCAACAGTGAAGAGAGGAGTGTGCTCCCTGCGGAGTGAAGACACTC 1740
Db 1681 GAAAAAAGAGAGCAACAGTGAAGAGAGGAGTGTGCTCCCTGCGGAGTGAAGACACTC 1740

```

```

Qy 1741 GGAAGGAAGCAGCTTGTGAGAGAGCAGACGCGCTGTGAGGAGAGATCACAATTAAT 1800
Db 1741 GGAAGGAAGCAGCTTGTGAGAGAGCAGACGCGCTGTGAGGAGAGATCACAATTAAT 1800
Qy 1801 GCAGTAAAGCCAGAAAGATGCTGTCTCTCTGCGCTCAGCTGTGTATTAATGATAT 1860
Db 1801 GCAGTAAAGCCAGAAAGATGCTGTCTCTCTGCGCTCAGCTGTGTATTAATGATAT 1860
Qy 1861 CACCTAAGGGGTGAGCTCTGAGAGCCCGCTCCGCTTCTTCTGAGATAGCATCCCTGG 1920
Db 1861 CACCTAAGGGGTGAGCTCTGAGAGCCCGCTCCGCTTCTTCTGAGATAGCATCCCTGG 1920
Qy 1921 TGTCCAGAGCTGAGAGTGCAGCTTGTGTGATGATGATCAAGACACCGCTGACCAT 1980
Db 1921 TGTCCAGAGCTGAGAGTGCAGCTTGTGTGATGATGATCAAGACACCGCTGACCAT 1980
Qy 1981 CAGCGGAGAGCAGAGCCCATATGTCAGATGCTCTGTCTGCTGTGTCTCAATCCCTAT 2040
Db 1981 CAGCGGAGAGCAGAGCCCATATGTCAGATGCTCTGTCTGCTGTGTCTCAATCCCTAT 2040
Qy 2041 TGTGAGAGCTTCTGTGACCTGTCTTCCAAAGCTGTAAACCTCACTGTGAAAGTCACT 2100
Db 2041 TGTGAGAGCTTCTGTGACCTGTCTTCCAAAGCTGTAAACCTCACTGTGAAAGTCACT 2100
Qy 2101 TAATGATTAATCTTAAATCTGTCTTCTCACTCAGGCTCTGATGATTTGATTTCT 2160
Db 2101 TAATGATTAATCTTAAATCTGTCTTCTCACTCAGGCTCTGATGATTTGATTTCT 2160
Qy 2161 CTTCATCCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
Db 2161 CTTCATCCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
Qy 2221 CACATGATTTGTGAGCAACAGAGGCTGTATAGCTTCACTTTTCCATTTGACAGG 2280
Db 2221 CACATGATTTGTGAGCAACAGAGGCTGTATAGCTTCACTTTTCCATTTGACAGG 2280
Qy 2281 CCAGACATGAAAGCATCTCTTTTATATGATGTGAGAACTTTGCAAGTATGATTTG 2340
Db 2281 CCAGACATGAAAGCATCTCTTTTATATGATGTGAGAACTTTGCAAGTATGATTTG 2340
Qy 2341 CATCTGATTTTCAAGTGTACATTTATTTTGTGATGTGAGAGATTTTGTGATTTT 2400
Db 2341 CATCTGATTTTCAAGTGTACATTTATTTTGTGATGTGAGAGATTTTGTGATTTT 2400
Qy 2401 TCCATGTCGATTCACAGCTACACACCAACCAATGAACTTTCAGATTTGAGAGCAG 2460
Db 2401 TCCATGTCGATTCACAGCTACACACCAACCAATGAACTTTCAGATTTGAGAGCAG 2460
Qy 2461 ACTCTGCTTCAATAGGCCCAAGGTAAGTGTCAACCTAGAACACTGTCTGACCGC 2520
Db 2461 ACTCTGCTTCAATAGGCCCAAGGTAAGTGTCAACCTAGAACACTGTCTGACCGC 2520
Qy 2521 AGAAGCGCTGCTTGAAGCTTGTATTTCAATGATGATGATGATGATGATGATGAT 2580
Db 2521 AGAAGCGCTGCTTGAAGCTTGTATTTCAATGATGATGATGATGATGATGATGAT 2580
Qy 2581 GAATGTTAGACTTTAAGATCATATCTG 2610
Db 2581 GAATGTTAGACTTTAAGATCATATCTG 2610

```

```

RESULT 3
BC014489 2725 bp mRNA linear PRI 04-OCT-2003
LOCUS Homo sapiens deach associated transcription factor 1, complete cds.
DEFINITION Variant 2, mRNA (cDNA clone MGC:23216 IMAGE:4901057), complete cds.
ACCESSION BC014489
VERSION BC014489.1 GI:15680266
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Hominidae; Homo.

```


QY 766 CTTGCGAAGAACCGGGAACAGAGCCCACTGAGAGGCCCTTGAAGAGGATCCAGAGTGC 825
 DB 739 CTTGCGAAGAACCGGGAACAGAGCCCACTGAGAGGCCCTTGAAGAGGATCCAGAGTGC 798
 QY 826 CTTGCGAAGAACCGGGAAGAGGATCCCGCCGAGACTGTGGGCTCCGAGCCAGTAC 885
 DB 799 CTTGCGAAGAACCGGGAAGAGGATCCCGCCGAGACTGTGGGCTCCGAGCCAGTAC 858
 QY 886 ACTGTGAGGCGCTCTCCCGAGTAAAGAGAGCCCGAGAACGATCAGGGGGTGTGTCC 945
 DB 859 ACTGTGAGGCGCTCTCCCGAGTAAAGAGAGCCCGAGAACGATCAGGGGGTGTGTCC 918
 QY 946 CAGGCTGGGAAAGAGTAAAGAGAGTAAAGTGAAGGAAAGCGGCTCAGGACATCAAA 1005
 DB 919 CAGGCTGGGAAAGAGTAAAGAGAGTAAAGTGAAGGAAAGCGGCTCAGGACATCAAA 978
 QY 1006 GATGAGAGCCCTGAGAGCTTGGGCGGAGCCGAGCCCTGAATGTAGAGGTTACGACCAC 1065
 DB 979 GATGAGAGCCCTGAGAGCTTGGGCGGAGCCGAGCCCTGAATGTAGAGGTTACGACCAC 1038
 QY 1066 GCCCTGTATTTGATTTGCCCGCAGCTTCAACAAACAGGTTATATTTGCTGTGACCGC 1125
 DB 1039 GCCCTGTATTTGATTTGCCCGCAGCTTCAACAAACAGGTTATATTTGCTGTGACCGC 1098
 QY 1126 TGTGAGAAATGGTTTCAATGCGAGTGTGTGGGCACTTCTGAGGCTCGAGGGAGCTTTTG 1185
 DB 1099 TGTGAGAAATGGTTTCAATGCGAGTGTGTGGGCACTTCTGAGGCTCGAGGGAGCTTTTG 1158
 QY 1186 GAAAGAGTGGGGAAGACTATATCTGCCCAAACCTGACCAATCTGCAAGTGCAGAGTAG 1245
 DB 1159 GAAAGAGTGGGGAAGACTATATCTGCCCAAACCTGACCAATCTGCAAGTGCAGAGTAG 1218
 QY 1246 ACTCATTCAGAAACCGCAGATCAGAGAGTAAATGAGACCTGAGAGTCTGTAGTGC 1305
 DB 1219 ACTCATTCAGAAACCGCAGATCAGAGAGTAAATGAGAGCTGAGAGTCTGTAGTGC 1278
 QY 1306 ACCGATGTACAGATGATGAGAACATAGACAGAAAGTCTAGCGAAAGACCAAGGATTAAG 1365
 DB 1279 ACCGATGTACAGATGATGAGAACATAGAGAGTCTAGCGAAAGACCAAGGATTAAG 1338
 QY 1366 GGTAGAAATTTGAGAAAGCTGCAATCTCAAGTGCAGAAAGAACTCAAGATCTTCCAGCT 1425
 DB 1339 GGTAGAAATTTGAGAAAGCTGCAATCTCAAGTGCAGAAAGAACTCAAGATCTTCCAGCT 1398
 QY 1426 GTGATAGAGGCGCTGTGTCTCAAAATGTATTGGCCCGGTGCTGTACGTGGCGCAG 1485
 DB 1399 GTGATAGAGGCGCTGTGTCTCAAAATGTATTGGCCCGGTGCTGTACGTGGCGCAG 1458
 QY 1486 CCGGACTGGGTACTGCGAGTAACTGTATCTCAACACGCGCAGCGACCAATGAG 1545
 DB 1459 CCGGACTGGGTACTGCGAGTAACTGTATCTCAACACGCGCAGCGACCAATGAG 1518
 QY 1546 TTTCTAAGCTCAGGTAAAGAACAGAACCCAAAGCCTAAAGAAAGTGAAGTGAAGCCA 1605
 DB 1519 TTTCTAAGCTCAGGTAAAGAACAGAACCCAAAGCCTAAAGAAAGTGAAGTGAAGCCA 1578
 QY 1606 GAGAGCCCAAGTCTTCCGAATGCGGTGCTCAGGAGGTATTAAATCTTCTGTGCAC 1665
 DB 1579 GAGAGCCCAAGTCTTCCGAATGCGGTGCTCAGGAGGTATTAAATCTTCTGTGCAC 1638
 QY 1666 AAGAGACCAAGCTCCGAAAGAAAGAAAGAACACAGAGTAAAGAGGAGTGTGCTCCCTG 1725
 DB 1639 AAGAGACCAAGCTCCGAAAGAAAGAAAGAACACAGAGTAAAGAGGAGTGTGCTCCCTG 1698
 QY 1726 CGGAGTGAAGCACTCGGGAAGAGACAGCTTGTGAGAGCAGCAGCCGTGCTGTGGCGAGC 1785
 DB 1699 CGGAGTGAAGCACTCGGGAAGAGACAGCTTGTGAGAGCAGCAGCCGTGCTGTGGCGAGC 1758
 QY 1786 GATCAACAATTACATGCTGATTAAGCCAGAAAGACTGTCTCTCCGCTGACCTGTTG 1845
 DB 1759 GATCAACAATTACATGCTGATTAAGCCAGAAAGACTGTCTCTCCGCTGACCTGTTG 1818
 QY 1846 TATAAATGATATCACTACCTAGGCGGTGGCCCTCTGGAGACCCCTCCGCTTCTTGTGATA 1905

DB 1819 TATAAATGATATCACTACCTAGGCGGTGGCCCTCTGGAGACCCCTCCGCTTCTTGTGATA 1878
 QY 1906 GGCATCCCTCTGGGCTGTTCAGAGACTTGGAGATGACGCTTGTGTAAGTGAATACAGCA 1965
 DB 1879 GGCATCCCTCTGGGCTGTTCAGAGACTTGGAGATGACGCTTGTGTAAGTGAATACAGCA 1938
 QY 1966 CACCGGCTGACCAATCAGCGGGAACAGAGCCCATGTCCAGAGTGCCTCTCCCTG 2025
 DB 1939 CACCGGCTGACCAATCAGCGGGAACAGAGCCCATGTCCAGAGTGCCTCTCCCTG 1998
 QY 2026 TGTTCATCCCTAGTCTGTCAGAGCTTCTGTCACTGTGTTTCCAAAGCTGTAACCTCACT 2085
 DB 1999 TGTTCATCCCTAGTCTGTCAGAGCTTCTGTCACTGTGTTTCCAAAGCTGTAACCTCACT 2058
 QY 2086 GGTGAACCTTCACTTAATGATTAATTTCTTAAATCTGTTTCACTCCAGGCTGTGT 2145
 DB 2059 GGTGAACCTTCACTTAATGATTAATTTCTTAAATCTGTTTCACTCCAGGCTGTGT 2118
 QY 2146 AAGTATTTGATTTCTTCAATCCAGTCTGATTTGATGAGCCACATGCCCCGACGCCAC 2205
 DB 2119 AAGTATTTGATTTCTTCAATCCAGTCTGATTTGATGAGCCACATGCCCCGACGCCAC 2178
 QY 2206 ATCCACCCCTGTCTGTCACATGAGTGTCTGACAAACAGGCTGATPACGCTTCACTTTT 2265
 DB 2179 ATCCACCCCTGTCTGTCACATGAGTGTCTGACAAACAGGCTGATPACGCTTCACTTTT 2238
 QY 2266 CCACATTTGTCACCGGCGACACATGAAAGCATCTTTTATTTATTTGTTGGAAATCTT 2325
 DB 2239 CCACATTTGTCACCGGCGACACATGAAAGCATCTTTTATTTATTTGTTGGAAATCTT 2298
 QY 2326 TGCAGTATGATGTTGATCTGATTTTTCAGGTGATCAATTTATTTGACTGCGCAGATAG 2385
 DB 2299 TGCAGTATGATGTTGATCTGATTTTTCAGGTGATCAATTTATTTGACTGCGCAGATAG 2358
 QY 2386 GGAATTTTTTTTTTTTTTCCATGTCAGATTCACACGCTTACACCCACATGACATTCGCA 2445
 DB 2359 GGAATTTTTTTTTTTTTTCCATGTCAGATTCACACGCTTACACCCACATGACATTCGCA 2416
 QY 2446 ACTTGGAAGG-CACACATCTCTGCTTCAATGAGCCCAAGTAAAGTGAAGTGAACACTAGA 2504
 DB 2417 ACTTGGAAGG-CACACATCTCTGCTTCAATGAGCCCAAGTAAAGTGAAGTGAACACTAGA 2476
 QY 2505 ACACGTGCTGACCGCAGAGCGGTGCTTGAAGTGTGATTTCAATGATGACTGCGCTT 2564
 DB 2477 ACACGTGCTGACCGCAGAGCGGTGCTTGAAGTGTGATTTCAATGATGACTGCGCTT 2536
 QY 2565 CTTGGCCTGCTCTTGTGAATGTTTGAAGCTTTAAGATCATATCTGT 2610
 DB 2537 CTTGGCCTGCTCTTGTGAATGTTTGAAGCTTTAAGATCATATCTGT 2582
 RESULT 4
 BC004237 2767 bp mRNA linear PRI 12-NOV-2003
 LOCUS
 DEFINITION Homo sapiens death associated transcription factor 1, transcript variant 1, mRNA (cdna clone MGC:896 IMAGS:2959493), complete cds.
 ACCESSION BC004237
 VERSION BC004237.2 GI:33872756
 KEYWORDS
 SOURCE MGC.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 2767)
 Strausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buettow, K.H., Scheffer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ueclin, T.B., Toshynski, S., Carninci, P., Prange, C., Rana, S.S., Loquellano, N.A., Peters, G.J.,

ABRAMSON, R.D., MULLAHY, S.J., BOSAK, S.A., MCEWAN, P.J.,
MCEWAN, K.J., MALEK, J.A., GUNARATNE, P.H., RICHARDS, S.,
MORLEY, K.C., HALE, S., GARCIA, A.M., GAY, L.J., HOLYK, S.W.,
VALLIANT, D.K., MUZY, D.M., SODERGREN, E.J., LU, X., GIBBS, R.A.,
FAHEY, J., HEITON, E., KETTEMAN, M., MADAN, A., RODRIGUES, S.,
SANCHEZ, A., WHITING, M., MADAN, A., YOUNG, A.C., SHEVCHENKO, Y.,
BOUFFARD, G.G., BLAKESLEY, R.W., TOUCHMAN, J.W., GREEN, E.D.,
DICKSON, M.C., RODRIGUEZ, A.C., GRIMWOOD, J., SCHMUTZ, J., MYERS, R.M.,
BUTERFIELD, Y.S., KRZYWINSKI, M.I., SKALKA, U., SMITH, D.E.,
SCHNERCH, A., SCHEIN, J.B., JONES, S.J., and MARRA, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL MEDLINE 2238257
PUBMED 12477932
TITLE 2 (bases 1 to 2767)
AUTHORS Straubberg, R.
JOURNAL Direct Submission
SUBMITTED (01-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Aug 19, 2003 this sequence version replaced gi:13278980.
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadan@systemsbio.org
Anup Madan, Jessica Fahey, Erin Heiton, Mark Ketteman, Anuradha
Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRL Plate: 2 Row: 0 Column: 16
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 18375620.
FEATURES SOURCE
1..2767
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:896 IMAG:2959493"
/tissue_type="Colon, adenocarcinoma"
/clone_1lb="NIH MGC 15"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1..2767
/gene="DATP1"
/note="synonyms: DIO1, FLJ11265, KIA0333"
/db_xref="locusid:11083"
/db_xref="MIM:604140"
282..1970
/codon_start=1
/product="death associated transcription factor 1, isoform
a"
/protein_id="AAH04237.1"
/db_xref="gi:13278981"
/translation="MIDKIN domain: PRAIKPTSKPRKMTGFRRTIAKREGADAS
ADPLEPPPOOLGLSRSGRGPRTKRTGRGRLTARRGSRSPVLSDEGRPSAC
PATDLETSBSVBSASTRSQPSASTVAKRPASSEKRVKGDHDTSDSDSGLT
LKEIQRLRKRRQEPFRLPKGIQSLRKRRRGGPAPVGSASDVPAVLPSKQ
PENQGVVSQAGKDEPRLRSKGAODIKDEPGLGRPKPCBCTYDPAVALCTICRQ
HNHFMICCDRCSEWPHGDCVGISBARGLERNEDYICPCTILQVODETHSTAD
OOBAFMRGDAGDGTCTSIGTIBOKSSEDOGKGRIBKRNAGSGKKLKIPIVLEAP
GASCKTGCCGCAVADPSVCSNDCILGHAAATMKTLSGGRKGRPKRKKMKRKP
SLPKCAQAGIKISSVHKRPAPKPKETTYKKAIVVPASBALGKKAECSSPTPSASD
HNYNVAVKERTPAAPSPILYKCMYHGLVGLDPSRSFWALIFWACPGLGVAALC"
1092..1241

ORIGIN
Query Match 96.5%; Score 2517.8; DB 9; Length 2767;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2542; Conservative 0; Mismatches 2; Indels 3; Gaps 2;
/note="PHD; Region: PHD-finger. PHD folds into an
interleaved type of Zn-finger chelating 2 Zn ions in a
similar manner to that of the RING and FYVE domains"
/db_xref="CDD:pfam00628"
65 GGAGCTTACCTCCAGGGGAACGCCCTGATATCTGAGTTGTTGAAAATAGAACCTTG 124
DB |||||
82 GGAGCTTATTCACGGGAACGCCCTGATATCTGAGTTGTTGAAAATAGAACCTTG 141
DB |||||
125 TTACTGTTGAACAGTGGCTGACCAAGTGTGTTGAGCTTGCTGTCTTGGAACCC 184
DB |||||
142 TTACTGTTGAACAGTGGCTGACCAAGTGTGTTGAGCTTGCTGTCTTGGAACCC 201
DB |||||
185 AGAGGTTTGTCTGCTCAGGGTTTGTGTTGTTATTTAGATTTCAGGAAAAGTGTCAAG 244
DB |||||
202 AGAGGTTTGTCTGCTCAGGGTTTGTGTTGTTATTTAGATTTCAGGAAAAGTGTCAAG 261
DB |||||
245 CTTTCAAGTTTGGACAGATATGACCAAGGGAACCCGACCAATGAGAGGACCTTA 304
DB |||||
262 CTTTCAAGTTTGGACAGATATGACCAAGGGAACCCGACCAATGAGAGGACCTTA 321
DB |||||
305 AGGACATCAAAACCCACAGCAAAAGATTCAAGAAAACATGAGGTTTTCAGAGACCTTA 364
DB |||||
322 AGGACATCAAAACCCACAGCAAAAGATTCAAGAAAACATGAGGTTTTCAGAGACCTTA 381
DB |||||
365 TCGCCAGAGAGAGGCGCAGGAGGACGCGAGAGCTGACCCACTGAGCCGCCACCCCTAC 424
DB |||||
382 TCGCCAGAGAGAGGCGCAGGAGGACGCGAGAGCTGACCCACTGAGCCGCCACCCCTAC 441
DB |||||
425 AGCAGCAGCTGGGCTCTGTCTCCGCGCGCAGTGGAGAGGCGCAAGCGCACTGAGCGG 484
DB |||||
442 AGCAGCAGCTGGGCTCTGTCTCCGCGCGCAGTGGAGAGGCGCAAGCGCACTGAGCGG 501
DB |||||
485 TGGAGCAGTCTCTGACCAATTTGCGCGCGCGCGAGAGAGAGCATCTGTCTCTCTG 544
DB |||||
502 TGGAGCAGTCTCTGACCAATTTGCGCGCGCGCGAGAGAGAGCATCTGTCTCTCTG 561
DB |||||
545 AGCATTTCTGTGAGGCTCCACGCTCTGCGCCGCAAGACGCCCAAGACGCTCCGAGGCA 604
DB |||||
562 AGCATTTCTGTGAGGCTCCACGCTCTGCGCCGCAAGACGCCCAAGACGCTCCGAGGCA 621
DB |||||
605 GCGTGAAGAGCGCTTCTGAGACCAAGGCGCCCAAGTCTCTTCCACAGCTGTGAAG 664
DB |||||
622 GCGTGAAGAGCGCTTCTGAGACCAAGGCGCCCAAGTCTCTTCCACAGCTGTGAAG 681
DB |||||
665 AACGACAGACCTCTCTGAAAGGTGAAGAGGGGATGACCAAGTGAACACTCCGATTA 724
DB |||||
682 AACGACAGACCTCTCTGAAAGGTGAAGAGGGGATGACCAAGTGAACACTCCGATTA 741
DB |||||
725 GTGACAGCATGAGCTCTGAAAGAGCTTCAAGATGCTTTCGAGAGAGCGGAGAC 784
DB |||||
742 GTGACAGCATGAGCTCTGAAAGAGCTTCAAGATGCTTTCGAGAGAGCGGAGAC 801
DB |||||
785 AGGAGGCCCACTGAGAGGCGCCCTGAAAGGATTCAGAGTGCCTGCGGAGAAAGCGCGG 844
DB |||||
802 AGGAGGCCCACTGAGAGGCGCCCTGAAAGGATTCAGAGTGCCTGCGGAGAAAGCGCGG 861
DB |||||
845 AGGAGGTCCTCCGCGAGACTGTGGGCTCCGAGGCGAGTGAACCTGTGAGGCGCTCTG 904
DB |||||
862 AGGAGGTCCTCCGCGAGACTGTGGGCTCCGAGGCGAGTGAACCTGTGAGGCGCTCTG 921
DB |||||
905 CCAAGTAAGCAGAGCCCGAGAAAGATCAAGAGGAGTGTGTCTCCAGGCTGGAGAAAGATGACA 964
DB |||||
922 CCAAGTAAGCAGAGCCCGAGAAAGATCAAGAGGAGTGTGTCTCCAGGCTGGAGAAAGATGACA 981
DB |||||
965 GAGAGAGTAAGTTGAGAGGAAAGCGGCTCAAGACATCAAAAGTAGAGGCTTGAGAGACT 1024
DB |||||
982 GAGAGAGTAAGTTGAGAGGAAAGCGGCTCAAGACATCAAAAGTAGAGGCTTGAGAGACT 1041
DB |||||

misc_feature

1025 TGGCCGACCGAAGCTGATGTGAGGGTTAAGACCCCAAGCCCTGTATTGCAATTTCC 1084
 1042 TGGCCGACCGAAGCTGATGTGAGGGTTAAGACCCCAAGCCCTGTATTGCAATTTCC 1101
 1085 GCCAGCTCACAACAGAGTTTATGATTTGCTGTACCGCTGTGAGAGATGGTTTCATG 1144
 1102 GCCAGCTCACAACAGAGTTTATGATTTGCTGTACCGCTGTGAGAGATGGTTTCATG 1161
 1145 GCGATTGTGTGGGCAATTTCTAGAGCTGAGGAGGCTTTTGGAAAGGAATGGGAAAGCT 1204
 1162 GCGATTGTGTGGGCAATTTCTAGAGCTGAGGAGGCTTTTGGAAAGGAATGGGAAAGCT 1221
 1205 ATATCTGCCCCAAGCTGACCAATTTCTGCAAGTGTGAGATGAGATCTATTCAGAAACGGAG 1264
 1222 ATATCTGCCCCAAGCTGACCAATTTCTGCAAGTGTGAGATGAGATCTATTCAGAAACGGAG 1281
 1265 ATCAGCAGGAAGCTAAATGAGAGCTGTGAGATGCTGATGAGCAACGATTTGTAACAATATAG 1324
 1282 ATCAGCAGGAAGCTAAATGAGAGCTGTGAGATGCTGATGAGCAACGATTTGTAACAATATAG 1341
 1325 GAACAAATGAGACAGAACTGAGCGAAGCAAGGATTAAGGGTATGATGAGAAAGCTG 1384
 1342 GAACAAATGAGACAGAACTGAGCGAAGCAAGGATTAAGGGTATGATGAGAAAGCTG 1401
 1385 CAAATCCAGTGGCAAGAAAGAACTCAAGATCTTCCAGCTGTGATAGAGGCGCTGTG 1444
 1402 CAAATCCAGTGGCAAGAAAGAACTCAAGATCTTCCAGCTGTGATAGAGGCGCTGTG 1461
 1445 CCTCAAAATGATTTGGCCCGGGTGTCTGTCAAGTGGCGAGCCGACTCGGTGATCTGCA 1504
 1462 CCTCAAAATGATTTGGCCCGGGTGTCTGTCAAGTGGCGAGCCGACTCGGTGATCTGCA 1521
 1505 GTATGACTGTATCTCTCAAAACGCGCGGAGGACATTAAGTTTCTTAAGCTCAGTTAAG 1564
 1522 GTATGACTGTATCTCTCAAAACGCGCGGAGGACATTAAGTTTCTTAAGCTCAGTTAAG 1581
 1565 AACAGAGCCCAAGCTTAAAGAAAGATGAGATGAGACGAGAAAGCCAGTCTTCCGA 1624
 1582 AACAGAGCCCAAGCTTAAAGAAAGATGAGATGAGACGAGAAAGCCAGTCTTCCGA 1641
 1625 AATCGGCTGTCTGAGGAGGATTAATATCTCTTGTGTGACAAAGACAGCTTCCAGAA 1684
 1642 AATCGGCTGTCTGAGGAGGATTAATATCTCTTGTGTGACAAAGACAGCTTCCAGAA 1701
 1685 AAAAAAGACCAAGTGAAGAGGACAGTGTGTCTCTGCGCGAGAGTAAACATCTCGGGA 1744
 1702 AAAAAAGACCAAGTGAAGAGGACAGTGTGTCTCTGCGCGAGAGTAAACATCTCGGGA 1761
 1745 AGGAAGCAGCTTGTGAGAGCAGACGCGCTGTGGCGAGGATCACAATTAACATGAG 1804
 1762 AGGAAGCAGCTTGTGAGAGCAGACGCGCTGTGGCGAGGATCACAATTAACATGAG 1821
 1805 TAAAGCAGAAAGAGCTGTGTCTCTGCGCGTCACTGTGTATTAATATGATATCAAC 1864
 1822 TAAAGCAGAAAGAGCTGTGTCTCTGCGCGTCACTGTGTATTAATATGATATCAAC 1881
 1865 TAGGGGTGTGCTCTGAGACCCCTCCGCTCTTCTGATGATGAGATCCCTCGGCGCTGT 1924
 1882 TAGGGGTGTGCTCTGAGACCCCTCCGCTCTTCTGATGATGAGATCCCTCGGCGCTGT 1941
 1925 CAGGACTGAGAGTTGACAGCTTGTGTTAAGCTGATCAAGACACCGGCTGACCATGAGC 1984
 1942 CAGGACTGAGAGTTGACAGCTTGTGTTAAGCTGATCAAGACACCGGCTGACCATGAGC 2001
 1985 GGAAGCAGAGCCCATGTCCAGAGTCCCTCTGCTGCTGTGTTCATCCCTAGTCTGT 2044
 2002 GGAAGCAGAGCCCATGTCCAGAGTCCCTCTGCTGCTGTGTTCATCCCTAGTCTGT 2061
 2045 AGGACTTCCTGTCTCTGTTTCCAAAGCTGTAAACCTCACTGTGAAAGTTCACCTTAAT 2104
 2062 AGGACTTCCTGTCTCTGTTTCCAAAGCTGTAAACCTCACTGTGAAAGTTCACCTTAAT 2121

2105 GATTGATCTTAAATCTCTGTGTTTCACTCTGAGGCTCTGATGATTTGATATCTCTTC 2164
 2122 GATTGATCTTAAATCTCTGTGTTTCACTCTGAGGCTCTGATGATTTGATATCTCTTC 2181
 2165 ATCCAGTCTGATGATGACCACTGCTCCGCGACCGCAATCAACCCCTGTCTGACCA 2224
 2182 ATCCAGTCTGATGATGACCACTGCTCCGCGACCGCAATCAACCCCTGTCTGACCA 2241
 2225 TGAGTTGTTCTGACCAAGGCTGTATAGCTTCACTTTTCCATTTGTCACAGGCGCAG 2284
 2242 TGAGTTGTTCTGACCAAGGCTGTATAGCTTCACTTTTCCATTTGTCACAGGCGCAG 2301
 2285 CACATGAAGCATCACTCTTTTATGTTGAGGGAATCTTTGCAAGTATGATGATC 2344
 2302 CACATGAAGCATCACTCTTTTATGTTGAGGGAATCTTTGCAAGTATGATGATC 2361
 2345 TGATTTGAGGTGATCAATTAATTTTGACTGGGAGATAGGGGATTTTTTTTTTCCA 2404
 2362 TGATTTGAGGTGATCAATTAATTTTGACTGGGAGATAGGGGATTTTTTTTTTCCA 2419
 2405 TGTCGATTTCAAGCTTCAACACCCCAATGAAACATTCGAACTCGAAGG-CACACACT 2463
 2420 TGTCGATTTCAAGCTTCAACACCCCAATGAAACATTCGAACTCGAAGGCAACACT 2479
 2464 CCTGCTTCAATGAGCCCAAGGATGATGATCAACCTTGAACACTCTGACCGCAGG 2523
 2480 CCTGCTTCAATGAGCCCAAGGATGATGATCAACCTTGAACACTCTGACCGCAGG 2539
 2524 ACGGTGCTTTGAGCTTGTATTTCTACATGATGATGATGATGATGATGATGATGAT 2583
 2540 ACGGTGCTTTGAGCTTGTATTTCTACATGATGATGATGATGATGATGATGATGAT 2599
 2584 TGTTAGACTCTTAAGATCATATCTG 2610
 2600 TGTTAGACTCTTAAGATCATATCTG 2626

RESULT 5
 AX878224
 LOCUS
 DEFINITION Sequence 13129 from Patent EP1074617.
 ACCESSION AX878224
 VERSION AX878224.1 GI:40032960
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 REFERENCE
 Ota,T., Isegai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
 Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
 TITLE Primers for synthesizing full-length cDNA and their use
 JOURNAL Patent: EP 1074617-A 13129 07-FEB-2001;
 Research Association for Biotechnology (JP)
 FEATURES
 source
 1..2772
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 301..1989
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAE90061.1"
 /db_xref="GI:40032961"
 /translation="MDPKSPNSNEAKAIKPSKBRKXWGPRBTITAKREGADAE
 ADLEPPPPQQLSLRSRGKRPKRTREYQFTITARRRRKSPSLSESGPTSC
 PADTARTASGVSASSTRSGPQASATAVAREPASSEKVGAGDHDHDSDSGLT
 LKELQNTLRREKREQEPTEPLKGIQSLRKRREBGAFTVGSASDIVEGLSKOE
 PENDDGVVSOAKODRPSKESKLEKAAQOIKEEPDLRKPKEGEGYPNALYICROP
 HNNRPMTCRCRCEWTFHGDVIGISSEARGLERNGEYICNCTILQVDETHSLAD
 QGEAKRPRPADGDTCSIGTIRBKSSEBDGIRKIRKRNPSGKKLKIIPQYIEAP
 GASKICIGPCCHVAQPDVSVCSNDCLIAHAIATKFTLSGREGQPKRKRRKMPKPK
 SLPCGAQGIKISVHKRPAPEKRETTVKAIVVPAVRSALGKEALCESSTPMSWASD

QY 2165 ATCCAGTCTGATTCATAGCCACATCCGACGCGCATCCATCCACCCCTGTCTGCACA 2224
 DB 2201 ATCCAGTCTGATTCATAGCCACATCCGACGCGCATCCATCCACCCCTGTCTGCACA 2260
 QY 2225 TGAAGTGTCTGACACACGCGCTGTATACGCTTCAATTTTCCATATGTCACGCGCAG 2284
 DB 2261 TGAAGTGTCTGACACACGCGCTGTATACGCTTCAATTTTCCATATGTCACGCGCAG 2320
 QY 2285 CACATGAAGATCACTCTTTTATGTGTGGGAATCTTGAAGTATGTTGCATC 2344
 DB 2321 CACATGAAGATCACTCTTTTATGTGTGGGAATCTTGAAGTATGTTGCATC 2380
 QY 2345 TGAATTTCAAGGTATCAATTTATTTTGAAGTGGGAGATTTTATTTTTCAC 2404
 DB 2381 TGAATTTCAAGGTATCAATTTATTTTGAAGTGGGAGATTTTATTTTTCAC 2438
 QY 2405 TGTTCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAAG-CAACACT 2463
 DB 2439 TGTTCGATTCACACGCTACACACCCACATGAACACATTCGAACTTCGAAAGCACAACACT 2498
 QY 2464 CCGCTTCTATAGGCGCCACCGGTAGTGTGATTCACACCTGAAACACTGCTCCGACCGCAG 2523
 DB 2499 CCGCTTCTATAGGCGCCACCGGTAGTGTGATTCACACCTGAAACACTGCTCCGACCGCAG 2558
 QY 2524 ACCGCTCCTTGAAGCTTGTATTTCTAATGATGATGCTGCTTCTTGCCCTCGTCTTGAA 2583
 DB 2559 ACCGCTCCTTGAAGCTTGTATTTCTAATGATGATGCTGCTTCTTGCCCTCGTCTTGAA 2618
 QY 2584 TGTTTAGACTCTTAAGATCATATCTTG 2610
 DB 2619 TGTTTAGACTCTTAAGATCATATCTTG 2645

RESULT 6
 LOCUS BD157094 2772 bp DNA linear PAT 17-JAN-2003
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD157094
 VERSION BD157094.1 GI:27862852
 KEYWORDS JP 2002191363-A/11937.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 TITLE 1 (Bases 1 to 2772)
 JOURNAL Oca, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J., Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
 Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002191363-A 11937 09-JUL-2002;
 HELIX RESEARCH INSTITUTE

COMMENT
 OS Homo sapiens (human)
 PN JP 2002191363-A/11937
 PD 09-JUL-2002
 PF 28-JUL-2000 JP 200028090
 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
 PI SAITO, YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
 PI JUNICHI NAGAI, TETSUJI OTSUKI
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
 10', C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT CDS Location/Qualifiers
 (301)..(1986).

FEATURES
 source 1..2772
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN

Query Match 96.4%; Score 2516.2; DB 6; Length 2772;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2541; Conservative 0; Mismatches 3; Indels 3; Gaps 2;
 QY 65 GGAGCTTACTCCAGGGAAAGAGGCTAGATATCTGATGTTGAAATATGAAGCTTG 124
 DB 101 GGAGCTTACTCCAGGGAAAGAGGCTAGATATCTGATGTTGAAATATGAAGCTTG 160
 QY 125 TTAATCGGAACAGTGGCTGACAAAGTGTGTTGTGAGCTGGCTGTCTCTGAGCCC 184
 DB 161 TTAATCGGAACAGTGGCTGACAAAGTGTGTTGTGAGCTGGCTGTCTCTGAGCCC 220
 QY 185 AGAGGTTTGTCTGCGCAGGGTTTGTGTTGTTATTTAGATTTCAAGGAAAGTGTCAAG 244
 DB 221 AGAGGTTTGTCTGCGCAGGGTTTGTGTTGTTATTTAGATTTCAAGGAAAGTGTCAAG 280
 QY 245 CTTTCAATGTTGAGCAGATATGACAAAGGCGACCGAGGAATAGAGAGGCACTTA 304
 DB 281 CTTTCAATGTTGAGCAGATATGACAAAGGCGACCGAGGAATAGAGAGGCACTTA 340
 QY 305 AGGCATCAAAACCAACAGCAAAAGTTCAAGAAACATGAGGTTTGAAGACCACTA 364
 DB 341 AGGCATCAAAACCAACAGCAAAAGTTCAAGAAACATGAGGTTTGAAGACCACTA 400
 QY 365 TCGCAAGCGAGAGGCGCAGGGGACCGGAGGCTGATCCCTGAGGCGCCACCCAC 424
 DB 401 TCGCAAGCGAGAGGCGCAGGGGACCGGAGGCTGATCCCTGAGGCGCCACCCAC 460
 QY 425 AGCAGCAGCTGGGCTGTCCCTGGGGGCAATGAGGAGCAGCCCAAGGCGCACTGAGCGG 484
 DB 461 AGCAGCAGCTGGGCTGTCCCTGGGGGCAATGAGGAGCAGCCCAAGGCGCACTGAGCGG 520
 QY 485 TGAGCAGTTCTGATCCATTCGCGCGCGCGCGCAGAGAGAGCATGCTGTCTCTG 544
 DB 521 TGAGCAGTTCTGATCCATTCGCGCGCGCGCGCAGAGAGAGCATGCTGTCTCTG 580
 QY 545 AGGATTTGTGTGAGGCCAGCTGCTGCCCCGCGCAGAGGCCGAGACACTCTCCAGGGCA 604
 DB 581 AGGATTTGTGTGAGGCCAGCTGCTGCCCCGCGCAGAGGCCGAGACACTCTCCAGGGCA 640
 QY 605 GCGTGAAGAGGCTTCTGAGACCAAGAGGAGCCCGCAGTCTGCTTCCAGCTGTGAAGG 664
 DB 641 GCGTGAAGAGGCTTCTGAGACCAAGAGGAGCCCGCAGTCTGCTTCCAGCTGTGAAGG 700
 QY 665 AACGACAGCTCTTCTGAAAAGGTGAAGAGGAGTGAACACGATGACACTTCCGATA 724
 DB 701 AACGACAGCTCTTCTGAAAAGGTGAAGAGGAGTGAACACGATGACACTTCCGATA 760
 QY 725 GTGACAGGATGCTGACCTTGAAGAAGCTTCAAGATGCTCTTGCAGAGAGCGGGAAC 784
 DB 761 GTGACAGGATGCTGACCTTGAAGAAGCTTCAAGATGCTCTTGCAGAGAGCGGGAAC 820
 QY 785 AGGAGCCCACTGAGAGGCTTGAAGAAGGATCCAGAGTGTCTGGGGAAGAGCGCGGG 844
 DB 821 AGGAGCCCACTGAGAGGCTTGAAGAAGGATCCAGAGTGTCTGGGGAAGAGCGCGGG 880
 QY 845 AGGAGGCTCCCGCGAGACTGTGGGCTCCGAGGCCAGTGAACACTGTGAGAGGCGTCTGC 904
 DB 881 AGGAGGCTCCCGCGAGACTGTGGGCTCCGAGGCCAGTGAACACTGTGAGAGGCGTCTGC 940
 QY 905 CCAATTAACAGAGACCCGAGAACATCAAGGAGGTTGTCTCCAGGCTGGGAAAGTGA 964
 DB 941 CCAATTAACAGAGACCCGAGAACATCAAGGAGGTTGTCTCCAGGCTGGGAAAGTGA 1000
 QY 965 GAGAGATTAATTTGAGAGGAAAGCGGCTCAGAGCATCAAGATGAGAGGCTTGAGACT 1024
 DB 1001 GAGAGATTAATTTGAGAGGAAAGCGGCTCAGAGCATCAAGATGAGAGGCTTGAGACT 1060
 QY 1025 TGGGCGGACCGAAGGCTGAATGTGAGGTTACAGACCCCAAGCGCTGTATTGCAATTTGCC 1084
 DB 1061 TGGGCGGACCGAAGGCTGAATGTGAGGTTACAGACCCCAAGCGCTGTATTGCAATTTGCC 1120
 QY 1085 GCCAGCTTCACAAACAGGTTTATGATTTGCTGTACCCGCTGTGAAAGATGTTCAATG 1144

Db	1121	GCACGCTTCAACAACAGGTTTATGATTGCTGTGACCGCTGTGAGAAATGGTTTCAATG	1180
Qy	1145	GCGATTTGTGTGGCAATTTCTGAGGCTCAAGGAGGCTTTTGGAAAGAAATGGGAAAGACT	1204
Db	1181	GCGATTTGTGTGGCAATTTCTGAGGCTGTGAGGAGGCTTTTGGAAAGAAATGGGAAAGACT	1240
Qy	1205	ATATTTGCCCAATCTGCACCAATTTCTGCAAGTGTGAGATGAGACTCTATTCAAGAAAGCGGAG	1264
Db	1241	ATATTTGCCCAATCTGCACCAATTTCTGCAAGTGTGAGAGAGAGACTCTATTCAAGAAAGCGGAG	1300
Qy	1265	ATCGACAGGAAGCTTAAATGAGAGCCTGTGAGATGCTGATGAGGACCGAATTGTACAAATATAG	1324
Db	1301	ATCGACAGGAAGCTTAAATGAGAGCCTGTGAGATGCTGATGAGGACCGAATTGTACAAATATAG	1360
Qy	1325	GAACAAATAGACAGAAAGTCTAGACGAGAGCAAGGGATTAAGGGTAAATTTAGAGAAAGCTG	1384
Db	1361	GAACAAATAGACAGAAAGTCTAGACGAGAGCAAGGGATTAAGGGTAAATTTAGAGAAAGCTG	1420
Qy	1385	CAATTCCAATGTGGCAAGAGAAATCTCAAGATCTTTCAGCTGTGATAGAGCGCTGTGTG	1444
Db	1421	CAATTCCAATGTGGCAAGAGAAATCTCAAGATCTTTCAGCTGTGATAGAGCGCTGTGTG	1480
Qy	1445	CCTCAAAATGTATTTGGGCCCGGGGTCTGTCAAGTGGGCGACGCGGACCTGGGTATCTGCA	1504
Db	1481	CCTCAAAATGTATTTGGGCCCGGGGTCTGTCAAGTGGGCGACGCGGACCTGGGTATCTGCA	1540
Qy	1505	GTAATGACTGTATCTCTCAAAACACGCGGAGGAGCAATAGAGTTTCTAAGCTCAGGTAAAG	1564
Db	1541	GTAATGACTGTATCTCTCAAAACACGCGGAGGAGCAATAGAGTTTCTAAGCTCAGGTAAAG	1600
Qy	1565	AACAGAAAGCCAAAGCTTAAAGAAAGATGAAGATGAAGCCAGAGAGCCCAAGTCTTCCGA	1624
Db	1601	AACAGAAAGCCAAAGCTTAAAGAAAGATGAAGATGAAGCCAGAGAGCCCAAGTCTTCCGA	1660
Qy	1625	AATCGGTGTCTCAGGCAAGGTATTAATAATCTCTCTGTGTGCAAGAAGCAAGCTCCAGAA	1684
Db	1661	AATCGGTGTCTCAGGCAAGGTATTAATAATCTCTCTGTGTGCAAGAAGCAAGCTCCAGAA	1720
Qy	1685	AAAAAGAGACCAACAGTGAAGAGGCAAGTGTGTCTCCGCGCGAGTGAAGCACTCGGGA	1744
Db	1721	AAAAAGAGACCAACAGTGAAGAGGCAAGTGTGTCTCCGCGCGAGTGAAGCACTCGGGA	1780
Qy	1745	AGGAAGCAGCTTGTGAGAGCAGCAAGCGGTGTGTGGGCGAGCGATCACAATTAACAATCAG	1804
Db	1781	AGGAAGCAGCTTGTGAGAGCAGCAAGCGGTGTGTGGGCGAGCGATCACAATTAACAATCAG	1840
Qy	1805	TAAAGCCAGAAAAGACTGTGTGTCTCCCTGTGGCCGTCACTGTATTATAAATGTAATGATCAC	1864
Db	1841	TAAAGCCAGAAAAGACTGTGTGTCTCCCTGTGGCCGTCACTGTATTATAAATGTAATGATCAC	1900
Qy	1865	TAGGGGTGTGGCTCCTGTGAGACCCCTCCCTTCTTCTGTGATAGCAATCCCTGTGGCTGTG	1924
Db	1901	TAGGGGTGTGGCTCCTGTGAGACCCCTCCCTTCTTCTGTGATAGCAATCCCTGTGGCTGTG	1960
Qy	1925	CAGGACTGGAGTTGACGCTTTGTGTGAAGCTGATCAGACACACCGGTGACCAATCAGC	1984
Db	1961	CAGGACTGGAGTTGACGCTTTGTGTGAAGCTGATCAGACACACCGGTGACCAATCAGC	2020
Qy	1985	GGGAAGCAGAGCCCAATGTTCAGAGATGCTCTCTGTGTGCCCTGTGTTCATCCCTAAGTGTG	2044
Db	2021	GGGAAGCAGAGCCCAATGTTCAGAGATGCTCTCTGTGTGCCCTGTGTTCATCCCTAAGTGTG	2080
Qy	2045	AGGACTTCTGTCACTGTTTTCCAAAGGCTGTAAACCTCACTGTGTAAAGCTTCACTTAAT	2104
Db	2081	AGGACTTCTGTCACTGTTTTCCAAAGGCTGTAAACCTCACTGTGTAAAGCTTCACTTAAT	2140
Qy	2105	GATTGATTTCTTAAATCTCTGTTTTCACTCTCAGGCTCTGTGAATTTGTATTTCTCTTC	2164
Db	2141	GATTGATTTCTTAAATCTCTGTTTTCACTCTCAGGCTCTGTGAATTTGTATTTCTCTTC	2200
Qy	2165	ATCCCACTGTGATTTGACCAACACTGTGCCGCGACGCGCAATCCACCTCTGTCTGACA	2224

Db	2201	ATCCGAGTGTGAATGATGATGACACACTGGCCCGGACGCGCACATCCACATCCACCCGTCGTGACA	2260
QY	2225	TGAGTTGTTCTGACACAAAGAGGGCTGTAAAGCTTCACAGTTTTTCCACATTGTGCAGGGCAG	2284
Db	2261	TGAGTTGTTCTGACAAAGAGGGCTGTAAAGCTTCACAGTTTTTCCACATTGTGCAGGGCAG	2320
QY	2285	CACATGAAGCATCACTTCCTTTTAAATGTGTGGGAATCTTTCGAAAGTATGATGTGCATC	2344
Db	2321	CACATGAAGCATCACTTCCTTTTAAATGTGTGGGAATCTTTCGAAAGTATGATGTGCATC	2360
QY	2345	TGATTTTTCAGGCTGTACATTTAATTTTGTGACTGGGCAATGAGGGGATTTTTTTTTTTTCCCA	2404
Db	2381	TGATTTTTCAGGCTGTACATTTAATTTTGTGACTGGGCAATGAGGGGATTTTTTTTTTTTCCCA	2438
QY	2405	TGTCCGATTTACACAGCTACACACCCACATGAAACATTTGAACTTTGGAAGG-CACACACT	2463
Db	2439	TGTCCGATTTACACAGCTACACACCCACATGAAACATTTGAACTTTGGAAGGCAACACACT	2498
QY	2464	CTCTCTTCATATAGGCCCAAGCGGTAAAGTAGTGTCAACCTTAGAACACTGTCTCTGACCGGAGG	2523
Db	2499	CTCTCTTCATATAGGCCCAAGCGGTAAAGTAGTGTCAACCTTAGAACACTGTCTCTGACCGGAGG	2558
QY	2524	ACGCGTGCCTTGGACCTTGTATTTCTACATGTGACTGTGCTTTCTTGCCCTGTGTCTTTGAA	2583
Db	2559	ACGCGTGCCTTGGACCTTGTATTTCTACATGTGACTGTGCTTTCTTGCCCTGTGTCTTTGAA	2618
QY	2584	TGTTTAGACTCTTAAGATCAADATCCCTG	2610
Db	2619	TGTTTAGACTCTTAAGATCAADATCCCTG	2645

RESULT 7	AC002127	2772 bp	mRNA	linear	PRI 01-AUG-2002
LOCUS	AK002127				
DEFINITION	Homo sapiens cDNA FLJ11265 fis, clone PLACE1009158.				
ACCESSION	AK002127				
VERSION	AK002127.1				
KEYWORDS	oligo capping; fis (full insert sequence).				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryotic Metazoa: Chordates; Craniates; Vertebrates; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS	1 Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H., Magatsuma, M., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K., Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahori, K., Maehub, Y., Ninomiya, K. and Iwayanagi, T.				
TITLE	NEO human cDNA sequencing project				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 2772)				
AUTHORS	Isogai, T. and Otsuki, T.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-FEB-2000) Takeo Isogai, Helix Research Institute, Genomics Laboratory / 1532-3 Yama, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@helix.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)				
COMMENT	NEO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.				
FEATURES	Location/Qualifiers				
source	1. 2772 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="PLACE1009158" /tissue_type="placenta" /clone_lib="PLACE1"				

CDS

/note="Cloning vector: pMB18F3"
301. 1989
/note="unamed protein product"
/codon_start=1
/protein_id="BAA92094.1"
/db_xref="GI:7021815"
/translation="MDKDGPSNEAPKAIKPTSEPRKTKTFRRTTIAKRBAGDAE
ADLEPPPOQIGLSIRSGROPKRTERVQFLTIARRGRSPVSLSDSEPTSC
PADLARTSSBSVPSASERTSGPOSASTAKERPASEKYGDHDHDSIDSLT
LKEIQNLRKRKRQSPTEPRPLKGIQSRILKRRREBPARTVGSASPTVGYLPSKBE
PENDOGVSOAGKDRESKLEGKAODIKDEPEDLAPRPECGYDNALYCIKRP
HNHRFMICDRCEWFPHGDCVGISSEARGILRENGEDYICPNCITLQVODLTHETAD
QOEAKMRPGADTDCTSIGTIBOKSSEDOGIKRIKRAAPSCKDLKIFQPIEAP
SLKCTGGCCVVAQPOSVCNNDCTIKMAATMKPISGKBOEKREKMRKEBP
SLKCCAGACIKISSVHRPAPEKRTTVKAVVVPASRLGEBACESTPEMAD
HNNAVPERKTAAPSPSLLYKCMTHLGVGLDPSRFTALPMACPLGVNALC"

ORIGIN

Query Match 96.4%; Score 2516.2; DB 9; Length 2772;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2541; Conservative 0; Mismatches 3; Indels 3; Gaps 2;

QY 65 GGAAGCTTACTCCAGGGAAAGAGCTTATGATATCTGAGTTGTGAAATAAGAGCCTG 124
DB 101 GGAGCTTACTCCAGGGAAAGAGCCTTATGATATCTGAGTTGTGAAATAAGAGCCTG 160
QY 125 TTACTCGTGAACAGTGGCTGACAAAGTGTGTTGTGAGCCTGCTGCTTGAAGCC 184
DB 161 TTACTCGTGAACAGTGGCTGACAAAGTGTGTTGTGAGCCTGCTGCTTGAAGCC 220
QY 185 AGAGGTTTCGCTGCGCAGGGTTTGTGTTGATTTAGATTCAAGGAAAGTGTCCAA 244
DB 221 AGAGGTTTCGCTGCGCAGGGTTTGTGTTGATTTAGATTCAAGGAAAGTGTCCAA 280
QY 245 CTTTCAGTGTGAGCAGGTATGACAAAGGCGACCCGAGCATGAGAGGACCTTA 304
DB 281 CTTTCAGTGTGAGCAGGTATGACAAAGGCGACCCGAGCATGAGAGGACCTTA 340
QY 305 AGGCCATCAAAAGCCAGCAGAAAGTTCAGGAAACATGGGGTTTTCAGAGACACTA 364
DB 341 AGGCCATCAAAAGCCAGCAGAAAGTTCAGGAAACATGGGGTTTTCAGAGACACTA 400
QY 365 TCGCAGAGCGAGAGGCGCAGAGGACGCGAGGCTGACCACTGAGCGCGACCCCA 424
DB 401 TCGCAGAGCGAGAGGCGCAGAGGACGCGAGGCTGACCACTGAGCGCGACCCCA 460
QY 425 AGCAGCAGCTGGGCTGTCCCTGCGCGCAGTGGGAGCAGCCCAAGGCGACTGAGCG 484
DB 461 AGCAGCAGCTGGGCTGTCCCTGCGCGCAGTGGGAGCAGCCCAAGGCGACTGAGCG 520
QY 485 TGAAGCAGTTCCTGACCAATGGCGCGCGCCGCGGAGAGAGCATGCTTCTTCCT 544
DB 521 TGAAGCAGTTCCTGACCAATGGCGCGCGCCGCGGAGAGAGCATGCTTCTTCCT 580
QY 545 AGGATTTCGTGAGAGCCCAAGTCCGCGCGCAGAGGCGGAGCAGCTCCGAGGGCA 604
DB 581 AGGATTTCGTGAGAGCCCAAGTCCGCGCGCAGAGGCGGAGCAGCTCCGAGGGCA 640
QY 605 GCGTGAAGAAAGCGCTTCTGAGACAGAGAGCGGCCCAAGTCTTCCACAGCTGTGA 664
DB 641 GCGTGAAGAAAGCGCTTCTGAGACAGAGAGCGGCCCAAGTCTTCCACAGCTGTGA 700
QY 665 AACGACAGCCTCTTCTGAAAGGTGAAGAGAGGAGATGACCAATGACACCTCCGAT 724
DB 701 AACGACAGCCTCTTCTGAAAGGTGAAGAGGAGATGACCAATGACACCTCCGAT 760
QY 725 GTGACAGGATGAGCCTTCTGAGAAAGAGCTTCAAGTCCGCTTTCAGAGAGCGGAA 784
DB 761 GTGACAGGATGAGCCTTCTGAGAAAGAGCTTCAAGTCCGCTTTCAGAGAGCGGAA 820
QY 785 AGGAGCCCACTGAGAGGCGCCCTGAAAGGAGATCCAGAGTCCGCTGCGAGAGAGCG 844
DB 821 AGGAGCCCACTGAGAGGCGCCCTGAAAGGAGATCCAGAGTCCGCTGCGAGAGAGCG 880

QY 845 AGAGGCTCCGCGCAGAGCTGTGGGCTCCGAGGCCAGTGAACATGTGAGGCGCTCTGC 904
DB 881 AGAGGCTCCGCGCAGAGCTGTGGGCTCCGAGGCCAGTGAACATGTGAGGCGCTCTGC 940
QY 905 CCAATTAACAGAGAGCCCGAGAACATAGAGGGGTTGTCTCCAGGCTGGGAAAGTACA 964
DB 941 CCAATTAACAGAGAGCCCGAGAACATAGAGGGGTTGTCTCCAGGCTGGGAAAGTACA 1000
QY 965 GAGGAGTAAAGTTGAGAGGAAAGCGCGCTCAGGACATCAAGATGAGAGCCTGAGACT 1024
DB 1001 GAGGAGTAAAGTTGAGAGGAAAGCGCGCTCAGGACATCAAGATGAGAGCCTGAGACT 1060
QY 1025 TGGCCGACAGAGCCTGAAATGTGAGGTTACAGACCCCAAGCCTGTATTCATTTGCC 1084
DB 1061 TGGCCGACAGAGCCTGAAATGTGAGGTTACAGACCCCAAGCCTGTATTCATTTGCC 1120
QY 1085 GCGAGCCTCAACAAACAGGTTATGATTTGCTGTGACCGCTGTGAAGATGTTCATG 1144
DB 1121 GCGAGCCTCAACAAACAGGTTATGATTTGCTGTGACCGCTGTGAAGATGTTCATG 1180
QY 1145 GCGATTGTGGGCAATTTCTGAGGCTGAGGAGGCTTTTGAAGAAAGAAAGTGAAG 1204
DB 1181 GCGATTGTGGGCAATTTCTGAGGCTGAGGAGGCTTTTGAAGAAAGAAAGTGAAG 1240
QY 1205 ATATCTGCCCCAACTGACACCATCTGCAAGTGCAGATGAGACTCATTCAGAAACGGAG 1264
DB 1241 ATATCTGCCCCAACTGACACCATCTGCAAGTGCAGATGAGACTCATTCAGAAACGGAG 1300
QY 1265 ATCAGCAGAGAGCTAAATGAGAGCTGAGAGTCTGATGAGCAAGATTGTACAGTATAG 1324
DB 1301 ATCAGCAGAGAGCTAAATGAGAGCTGAGAGTCTGATGAGCAAGATTGTACAGTATAG 1360
QY 1325 GAAACATGAGACAGAGTCTTACGAAACCAAGGATTAAGGTGATGAGAAAGCTG 1384
DB 1361 GAAACATGAGACAGAGTCTTACGAAACCAAGGATTAAGGTGATGAGAAAGCTG 1420
QY 1385 CAAATCCAGTGGCAAGAAAGAACTCAAGATCTTCCAGCTGTGTAGAGGCGCTGTGTG 1444
DB 1421 CAAATCCAGTGGCAAGAAAGAACTCAAGATCTTCCAGCTGTGTAGAGGCGCTGTGTG 1480
QY 1445 CCTCAAAATGTATTTGCCCCCGGAGTGTGTCAGTGGCGCAGCCGACCTCGGTATCTG 1504
DB 1481 CCTCAAAATGTATTTGCCCCCGGAGTGTGTCAGTGGCGCAGCCGACCTCGGTATCTG 1540
QY 1505 GTAATGACTGTATCTTCAAACACCGCGAGGAGCAATGAAGTTTCTAGGCTCAGGTAAG 1564
DB 1541 GTAATGACTGTATCTTCAAACACCGCGAGGAGCAATGAAGTTTCTAGGCTCAGGTAAG 1600
QY 1565 AACGAAAGCCAAAGCTTAAAGAAAGATGAAGTGAAGCCAGAGAGCCCAAGTCTTCGGA 1624
DB 1601 AACGAAAGCCAAAGCTTAAAGAAAGATGAAGTGAAGCCAGAGAGCCCAAGTCTTCGGA 1660
QY 1625 AATGCGTGTCTCAGGCAAGTATTAATAATCTCTTCTGTGCAACAGAGCAAGCTTCAGAA 1684
DB 1661 AATGCGTGTCTCAGGCAAGTATTAATAATCTCTTCTGTGCAACAGAGCAAGCTTCAGAA 1720
QY 1685 AAAAAAGACCAAGTGAAGAGGCAAGTGTGTCTCCGCGCGGAGTGAAGACTCCGGGA 1744
DB 1721 AAAAAAGACCAAGTGAAGAGGCAAGTGTGTCTCCGCGCGGAGTGAAGACTCCGGGA 1780
QY 1745 AGGAGGAGCTTGTGAGAGAGCAAGCGCGTGTGGGCGAGGATCACAATTCAATGACAG 1804
DB 1781 AGGAGGAGCTTGTGAGAGAGCAAGCGCGTGTGGGCGAGGATCACAATTCAATGACAG 1840
QY 1805 TAAAGCCAGAAAAGACTGCTGCTCCCTGCGCGTCACTGTGTATAATGTATATCAC 1864
DB 1841 TAAAGCCAGAAAAGACTGCTGCTCCCTGCGCGTCACTGTGTATAATGTATATCAC 1900
QY 1865 TAGGGGTTGGCTCTGTGACCCCTTCCGCTTCTTTCTGTGATAGCCATCCCTGGGCTGTG 1924
DB 1901 TAGGGGTTGGCTCTGTGACCCCTTCCGCTTCTTTCTGTGATAGCCATCCCTGGGCTGTG 1960

OY		1925	CAGGACTGGAGGTGCAGCTTGTTGAATCATCAACGCCCGTGCACCATTACG	1984
Dd		1961	CAGGACTGGAGGTGCAGCCTTGTTGAATCATCAACGCCCGTGCACCATTACG	2020
OY		1985	GGAAGAAGACAGCCCATGTCAGAGATGCCCTGCTGCCCCCTGTGTCATCCCTAATCTGTC	2044
Dd		2021	GGAAGAAGACAGCCCATGTCAGAGATGCCCTGCTGCCCCCTGTGTCATCCCTAATCTGTC	2088
OY		2045	AGGACTTCCTGTCACTGTTTTCCAAAGCTGTAACCTCACTGCTGAACGTTCACTTAAAT	2104
Dd		2081	AGGACTTCCTGTCACTGTTTTCCAAAGCTGTAACCTCACTGCTGAACGTTCACTTAAAT	2140
OY		2105	GATTGAATCTCTTAATCTGTGTTTTCACTCTCACAGGCTCGTAGAATTTGTAATCTCTTC	2166
Dd		2141	GATTGAATCTCTTAATCTGTGTTTTCACTCTCACAGGCTCGTAGAATTTGTAATCTCTTC	2200
OY		2165	ATCCCAAGCTGATATGCATAGCCACAACCTGCCGACACGCCACATCCACCCTGTCTGCACA	2224
Dd		2201	ATCCCAAGCTGATATGCATAGCCACAACCTGCCGACACGCCACATCCACCCTGTCTGCACA	2268
OY		2225	TGAGTGTGTTCTGCACAACAGCCCTGTATAAGCTTCAGTTTTTCCACATTTGTCCAAGCCGAC	2284
Dd		2261	TGAGTGTGTTCTGCACAACAGCCCTGTATAAGCTTCAGTTTTTCCACATTTGTCCAAGCCGAC	2320
OY		2285	CACATGGAAGAGATCACTCTTTTTATGTTGTTGGGAATCTTGTGAAGTTAGTTGTCATC	2344
Dd		2321	CACATGGAAGAGATCACTCTTTTTATGTTGTTGGGAATCTTGTGAAGTTAGTTGTCATC	2380
OY		2345	TGATTTTCAGGCTGACATTTATTTTGTGAATGGGCAGATAGGGGATTTTTTTTTTTTCCA	2404
Dd		2381	TGATTTTCAGGCTGACATTTATTTTGTGAATGGGCAGATAGGGGA- TTTTTTTTTTTTCCA	2438
OY		2405	TGTCGATTTCAACAGCTTACACACCCACATGAAACATTTGGAATCTTGAAG- CACACACT	2463
Dd		2439	TGTCGATTTCAACAGCTTACACACCCACATGAAACATTTGGAATCTTGAAGCACAACACT	2498
OY		2464	CTGTCTTATAGGCCCCCAAGGTAAGTGAAGTTCAACCTTGAACACTGTCTGTGACCCGACAG	2523
Dd		2499	CTGTCTTATAGGCCCCCAAGGTAAGTGAAGTTCAACCTTGAACACTGTCTGTGACCCGACAG	2558
OY		2524	ACGGGTGCTTTGGACTTGATATCTACACATGTAACCTGAGCTTCTCTTCTCTTGA	2583
Dd		2559	ACGGGTGCTTTGGACTTGATATCTACACATGTAACCTGAGCTTCTCTTCTCTTGA	2618
OY		2584	TGTTTAGACTCTTAGATCATATCTCTG 2610	
Dd		2619	TGTTTAGACTCTTAGATCATATCTCTG 2645	
RESULT 8				
BC000770				
LOCUS				
DEFINITION	Homo sapiens death associated transcription factor 1, transcript variant 3, mRNA (cDNA clone MGC:3257 IMAGE:3506207), complete cds.			
ACCESSION	BC000770			
VERSION	BC000770.2 GI:33990873			
KEYWORDS	MGC.			
SOURCE	Homo sapiens (human)			
ORGANISM	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE				
AUTHORS	Strauberg,R.L., Feingold,B.A., Grouse,L.H., Deryg,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shemen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Steele,T.E., Brownstein,M.J., Uedlin,T.B., Casavant,T.D., Canciani,P., Prange,C., Raha,S., Loquellano,N.A., Peters,G.U., Abramson,R.D., Mullahy,S.J., Boasak,S.A., McEwan,P.J., McEranan,K.U., Malek,U.A., Gunaratne,P.H., Richards,S., Motley,K.C., Hale,S., Garcia,A.M., Gay,L.U., Huliyil,S.W.,			

TITLE	JOURNAL	MEDLINE PUBMED	AUTHORS	REFERENCE
human and mouse cDNA sequences	Proc. Natl. Acad. Sci. U.S.A.	99 (26), 16899-16903 (2002)		
Generation and initial analysis of more than 15,000 full-length				
22386257				
12477932				
Strausberg, R.				
Direct Submission				
Submitted (15-NOV-2000) National Institutes of Health, Mammalian				
Gene Collection (MGC), Cancer Genomics Office, National Cancer				
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,				
USA				
NIH-MGC Project URL: http://mgc.nci.nih.gov				
On Aug 20, 2003 this sequence version replaced gi:12653952.				
Contact: MGC help desk				
Email: cgapbs-remail.nih.gov				
Tissue Procurement: DCTD/DTP				
cDNA Library Preparation: Rubin Laboratory				
DNA Sequencing by: Institute for Systems Biology				
http://www.systemsbio.org				
contact: amadan@systemsbiology.org				
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha				
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting				
Clone distribution: MGC clone distribution information can be found				
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov				
Series: IRAL Plate: 7 Row: 1 Column: 2				
This clone was selected for full length sequencing because it				
passed the following selection criteria: Hexamer frequency ORF				
analysis, similarity but not identity to protein.				
Location/Qualifiers				
1. 2407				
/organism="Homo sapiens"				
/mol_type="mRNA"				
/db_xref="taxon:9606"				
/clone="MGC:3257 IMAGE:3506207"				
/tissue_type="Kidney, renal cell adenocarcinoma"				
/clone_id="Nih MGC_14"				
/lab_host="DH10B-R"				
/note="Vector: POT87"				
1. 2407				
/gene="DATP1"				
/note="synonyms: DIOL, FLJ11265, KIAA0333"				
/db_xref="LocustID:"11083"				
/db_xref="MIM:604140"				
201..1835				
/codon_start=1				
/product="death associated transcription factor 1, isoform				
b"				
/protein_id="AAH00770.1"				
/db_xref="GI:12653953"				
/db_xref="locusid:11083"				
/translation="MDKDDPSBEAPKAIKPTSKPRKYGRRTIARBAAGABR				
ADLLEPPPOQGLSLKRSRGROPKATVERPQLTARRRRGRSPYSLADESEPTSC				
PADDAETSGVESASSTRSGPASSTAWKBPASSERYKGADDDHDSDSDXDKIG				
LKEILQRLLRKREORPTERPLKGIQSRLTKRBEBGPAIVYGSAITVGVLVSXGB				
PENDGVSVSAGKDDESKLGGKAODIKOEBSGDGLGRPECEGYDPNALYCICRP				
HNNRPVICDRCEWFGDCVQISBARGDLIRNGRDYICPNCTIILOVDETHSRAD				
OQAKMRPGHADDTCTSIGTIEQSSBDQIKRIKRNPSGKKTLKIROPVIAR				
GASKRGPGCCHVAQPDYSVCSMDCLTKANAATMKFLSSGEOKPYPKEMAKKPYP				
SLEPCGAAGIKIISKVHKRPAPERKETTVKCAVVAFARSBALGKAACESTSWASD				
HNNAVKPEKTAPSPSLLYCKSGKTYLSHPSLIA"				
1011..1160				
/note="PHD; Region: PHD-finger. PHD folds into an				
interleaved type of Zn-finger chelating 2 Zn ions in a				

similar manner to that of the RING and FYVE domains"
/db_xref=CDD:pfam00628"

Query Match 74.4%; Score 1941; DB 9; Length 2407;
Best Local Similarity 88.7%; Pred. No. 0;
Matches 2259; Conservative 0; Mismatches 0; Indels 288; Gaps 3;

65 GGAGCTTAATCTCAGGGAACAGCCTCTAGATATCTGAGTTGTTGAAATACGAAGCTG 124
1 GGAGCTTAATCTCAGGGAACAGCCTCTAGATATCTGAGTTGTTGAAATACGAAGCTG 60

125 TTATCTCGGAACAGTGGCTGCAACAGTGTGTGTGAGCTGAGCTGTCTGTGAGACC 184
61 TTATCTCGGAACAGTGGCTGCAACAGTGTGTGTGAGCTGAGCTGTCTGTGAGACC 120

185 AGAGGTTTCGTCTGCGCAGGGTTTGTGTTGATTTAGATTTCAGGGAAGTGTCCAAG 244
121 AGAGGTTTCGTCTGCGCAGGGTTTGTGTTGATTTAGATTTCAGGGAAGTGTCCAAG 180

245 CTTTCAGTGTGAGCAGATATGACGACAAAGGCAACCCGAGCAATGAGAGCACTTA 304
181 CTTTCAGTGTGAGCAGATATGACGACAAAGGCAACCCGAGCAATGAGAGCACTTA 240

305 AGGCATCAAAACCAACAGCAAAAGATTCAAGAAAACATGGGGTTTTCGAAAGACACTA 364
241 AGGCATCAAAACCAACAGCAAAAGATTCAAGAAAACATGGGGTTTTCGAAAGACACTA 300

365 TCGCCAAACGAGAGGAGCGCAGGGGACGCGAGGCTGACCACTGAGCCGCCCAAC 424
301 TCGCCAAACGAGAGGAGCGCAGGGGACGCGAGGCTGACCACTGAGCCGCCCAAC 360

425 AGCAGCAGTGGGCTGTTCCTTCGCGGCGCATGTGAGAGGCAAGCGCACTGAGCGCG 484
361 AGCAGCAGTGGGCTGTTCCTTCGCGGCGCATGTGAGAGGCAAGCGCACTGAGCGCG 420

485 TGGAGCAGTTCCTGACATTTGCGCGCGCGCGCGAGGAGAGCATGCTGTCTCTCGT 544
421 TGGAGCAGTTCCTGACATTTGCGCGCGCGCGCGAGGAGAGCATGCTGTCTCTCGT 480

545 AGGATTCGTGTAAGCCCAAGTCTGCGCGCGCGCAAGACCGCGAGACAGCTTCGAGGCA 604
481 AGGATTCGTGTAAGCCCAAGTCTGCGCGCGCGCGCAAGACCGCGAGACAGCTTCGAGGCA 540

605 GCGTGGAAAGGCGCTTCTGAGACCAAGAGCGCGCGCGCAATCTGCTTCCAGAGTGAAG 664
541 GCGTGGAAAGGCGCTTCTGAGACCAAGAGCGCGCGCGCAATCTGCTTCCAGAGTGAAG 600

665 AACGACCAAGCTTCTGAAAAAGTGAAGAGAGGAGATGACCAAGATGACACTCCGATA 724
601 AACGACCAAGCTTCTGAAAAAGTGAAGAGAGGAGATGACCAAGATGACACTCCGATA 660

725 GTGACAGGATGAGCTGACCTTGAAGAGCTTCAAGATTCGCTTCCAGAGAGCGGAAAC 784
661 GTGACAGGATGAGCTGACCTTGAAGAGCTTCAAGATTCGCTTCCAGAGAGCGGAAAC 720

785 AGGAGCCCACTGAGAGAGCGCTTGAAGAGGATTCAGAGTCCGCTGCGAGAGAGCGCGG 844
721 AGGAGCCCACTGAGAGAGCGCTTGAAGAGGATTCAGAGTCCGCTGCGAGAGAGCGCGG 780

845 AGGAGGATCCCGCCGAGAGCTGTGGGCTCCGAGGCAAGTGAACACTGTGAGAGGCGTCTGC 904
781 AGGAGGATCCCGCCGAGAGCTGTGGGCTCCGAGGCAAGTGAACACTGTGAGAGGCGTCTGC 840

905 CCAAGTAACAGAGAGCCGAGAGACGATCAGAGGATTTGTCTCCAGGCTGAGAAAGATGACA 964
841 CCAAGTAACAGAGAGCCGAGAGACGATCAGAGGATTTGTCTCCAGGCTGAGAAAGATGACA 900

965 GAGAGATTAATTTGAGAGGAAAGCGGCTCAGAGACATCAAGAAATGAGAGAGCTGAGAGCT 1024
901 GAGAGATTAATTTGAGAGGAAAGCGGCTCAGAGACATCAAGAAATGAGAGAGCTGAGAGCT 960

1025 TGGGCGGACCGAAGCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTTGCAATTGCGC 1084

||||| 961 TGGGCGGACCGAAGCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTTGCAATTGCGC 1020

1085 GCCAGCTTCAACAACAAGCTTTATGATTTTCTGTGACCCCTGTGTAAGAAATGCTTTACATG 1144

1021 GCCAGCTTCAACAACAAGCTTTATGATTTTCTGTGACCCCTGTGTAAGAAATGCTTTACATG 1080

1145 GCGATTGTGTGGGCAATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGAAATGGGGAAGACT 1204

1081 GCGATTGTGTGGGCAATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGAAATGGGGAAGACT 1140

1205 ATATCTGCCCAAACTGCAACAATTCGCAAGTGCAGAGATGAGACTCATTCAGAAAAGCGCAG 1264

1141 ATATCTGCCCAAACTGCAACAATTCGCAAGTGCAGAGATGAGACTCATTCAGAAAAGCGCAG 1200

1285 ATCCAGAGAAAGCTAAATGAGAACCTGAGAGATGCTGATGAGCAACGATTGTACAAGATATG 1324

1201 ATCCAGAGAAAGCTAAATGAGAACCTGAGAGATGCTGATGAGCAACGATTGTACAAGATATG 1260

1325 GAAACAATAGACAGAAAGTCTAGCGAAGACCAAGGGATTAAGGGTAAAGTTGAGAAAAGCTG 1384

1281 GAAACAATAGACAGAAAGTCTAGCGAAGACCAAGGGATTAAGGGTAAAGTTGAGAAAAGCTG 1320

1385 CAATTCAGATGGCAAGAAAGAACTCAAGATCTTCAAGCTGTGATGAGCGGCTGTG 1444

1321 CAATTCAGATGGCAAGAAAGAACTCAAGATCTTCAAGCTGTGATGAGCGGCTGTG 1380

1445 CCTCAAAATGTATTTGCGCGCGCGGCTGTGACAGTGGCGCGCACTCCGGTACTGCA 1504

1381 CCTCAAAATGTATTTGCGCGCGCGGCTGTGACAGTGGCGCGCACTCCGGTACTGCA 1440

1505 GTAAATGACTGTATCTCTCAAAACGCGCGCAGCAACAATGAAGTTTCTAAGCTCAGGTAAG 1564

1441 GTAAATGACTGTATCTCTCAAAACGCGCGCAGCAACAATGAAGTTTCTAAGCTCAGGTAAG 1500

1565 AACGAGACCAAAAGCTTAAAGAAAGATGAAGATGAAGCCAGAGAGCCCACTTTCCGA 1624

1501 AACGAGACCAAAAGCTTAAAGAAAGATGAAGATGAAGCCAGAGAGCCCACTTTCCGA 1560

1625 AATGGGTGCTCAGGCAAGATTAATAATCTTCTGTGCAACAAGACCAAGCTCCAGAAA 1684

1561 AATGGGTGCTCAGGCAAGATTAATAATCTTCTGTGCAACAAGACCAAGCTCCAGAAA 1620

1685 AAAAAGAGACCAACATGTAAGAAAGCAGTGTGTGCTTCCGCGAGTGAAGCACTTCGGGA 1744

1621 AAAAAGAGACCAACATGTAAGAAAGCAGTGTGTGCTTCCGCGAGTGAAGCACTTCGGGA 1680

1745 AGGAAGCAGCTTTGTGAGACAGACGCGGTCTGTGGGCGAGCGATCACAATTCAATGCAAG 1804

1681 AGGAAGCAGCTTTGTGAGACAGACGCGGTCTGTGGGCGAGCGATCACAATTCAATGCAAG 1740

1805 TAAAGCCAGAAAAGCTGTCTGCTCCCTGCGGTCACTGTGTAATAATGATATGATCAAC 1864

1741 TAAAGCCAGAAAAGCTGTCTGCTCCCTGCGGTCACTGTGTAATAATGATATGATCAAC 1788

1865 TAGGGGTTGACCTCTGAGACCCCTCCGTTCTTCTGATAGCAATCCCTCGGCGCTGTTC 1924

1789 ----- 1788

1925 CAGGACTGGAGATTGACAGCTTTGTGTAACTGATACAGACACCGGCTGCAACATCAAC 1984

1789 ----- 1788

1985 GGGAAAGCAGAGCCCATGTCCAGAGATGCTTCCTGCTGTGCTCATCCATGCTGTTC 2044

1789 ----- 1788

2045 AGGACTTCTGTCACTGTTTTCAAAGCTGTAACTCACTGTGTAAGCTTCACTTAAT 2104

1789 ----- 1788

2105 GATTGATTTCTTAATCTCTGTTTTCACCTCAAGGCTCTGTGTAAGTATTTGATTTCTTTC 2164

QY 740 TGAACCTTGAAGAGCTTCAGAAATGCGCTTGGCAGAGAGCGGAAACAGAGCCCACTGAGA 799
 DB 698 TTAGCTTGAAGAGAGCTTCAGAAACGCGCTTGGAGAAAGCAGAGAGAGAACTTGTGAGA 757
 QY 800 GGCCTCTGAAGAGAGATCCAGAGTCCCTTGGAGAAAGAGCGCGGAGAGGGGTCCCGCG 859
 DB 758 GGTCTCTGAAGAGAGATCCAGAGTCCCTTGGAGAAAGAGCGAGAGAGAAATTTCTGCG 817
 QY 860 AGACTGTGGCTCCGAGGCGCAGTACACTGTGAGAGGCGCTTCCGACATGAGAGAGC 919
 DB 818 AAATCTGAGAGTGTCCAAATAGGCGAGTCCGAGCAGAGCAGACTCTCTGTATGAGAGAGC 877
 QY 920 CCGAGAGAGATCAGAGAGGTTGTGTCCAGGCTGAGGAAAGATGACAGAGAGAGTGTG 979
 DB 878 CTGAGGCTGAGTACAGAGAGCAGATGTCCAGTACAGAGAGAGATGACATGAGAAATCAGTTGG 937
 QY 980 AGGAGAGAGCGGCTCAGAGCAGTCAAAAGATGAGAGAGCTGAGAGACTTGGCGCAGCAG 1039
 DB 938 AAGGAGAGGCGACTCAGAGAAATACAGAGAAATCCCGAGGAAAGCGGCAAAACCAAGC 997
 QY 1040 CTGAATGTAGAGGTTAGCAGACCCCAAGCGCTGTATTTGATTTGCCCGCAGCTTCAACA 1099
 DB 998 CTGAGTGTGAGGTTAGCAGACCCCAAGCGCTGTATTTGATTTGCCCGCAGCTTCAACA 1057
 QY 1100 AAGGTTTATGATTTGCTGTGAGCAGCTGTGAGAGATGTTTATGAGGAGTGTGTGGCA 1159
 DB 1058 AAGGTTTATGATTTGCTGTGAGCAGCTGTGAGAGATGTTTATGAGGAGTGTGTGGCA 1117
 QY 1160 TTTCTGAGAGCTCAGAGAGAGGCTTTTGGAGAAAGATGAGAGACTATATCTGCCCAACT 1219
 DB 1118 TTTCTGAGAGCTCAGAGAGAGGCTTTTGGAGAAAGATGAGAGACTATATCTGCCCAACT 1177
 QY 1220 GCACATTTCTGAGAGTGCAGAGATGAGAGTCACTTCAAGAAACCGCAGATGACAGAGAGTA 1279
 DB 1178 GCACATTTCTGAGAGTGCAGAGATGAGAGTCACTTCAAGAAACCGCAGATGACAGAGAGTA 1237
 QY 1280 AATGAGAGAGCTGAGAGATGAGAGTCACTTCAAGAAACCGCAGATGAGAGAGTA 1339
 DB 1238 GGTGCAATCTGTGGGTGTGATGAGAGTCACTTCAAGAAACCGCAGATGAGAGAGTA 1297
 QY 1340 AGTCTAGCAGAGAGCAGAGATGAGAGTCACTTCAAGAAACCGCAGATGAGAGAGTA 1399
 DB 1298 AGTCTAGCAGAGAGCAGAGATGAGAGTCACTTCAAGAAACCGCAGATGAGAGAGTA 1357
 QY 1400 AGAAGAAATCTCAGATCTTCCAGCTGTGATGAGAGGCGCTGTGCTCAAAATGTATTG 1459
 DB 1358 AGAAGAAATCTCAGATCTTCCAGCTGTGATGAGAGGCGCTGTGCTCAAAATGTATTG 1417
 QY 1460 GCGCCGGGTGTGCTCAGAGTGGCAGAGCGCAGCTGCTGTGATGAGAGTCACTTCAAAATG 1519
 DB 1418 GCGCCGGGTGTGCTCAGAGTGGCAGAGCGCAGCTGCTGTGATGAGAGTCACTTCAAAATG 1477
 QY 1520 TCAAAACGCGCAGCAGCAATGAGTTTCTAGCTCAGGTAAAGAACAGAGCCAAAGC 1579
 DB 1478 TCAAAACGCGCAGCAGCAATGAGTTTCTAGCTCAGGTAAAGAACAGAGCCAAAGC 1537
 QY 1580 CTAAAGAAAGATGAAATGAGAGAGAGAGCCAGCTTCCGAAATGCGCTCTCAGG 1639
 DB 1538 CCAAGAAAGATGAAATGAGAGAGAGAGCCAGCTTCCGAAATGCGCTCTCAGG 1597
 QY 1640 CAGGTATTTAAATCTCTTCTGTGCAAGAGAGAGCAGCTCCAGAAATGAGAGCAGCAG 1699
 DB 1598 TGGGATTTAAATCTCTTCTGTGCAAGAGAGAGCAGCTCCAGAAATGAGAGCAGCAG 1657
 QY 1700 TGAAGAGAGCAGTGTGTCTCTGCGCGAGTGAAGCCTCGGAGAGAGCAGCTTGTG 1759
 DB 1658 TGAAGAGAGCAGTGTGTCTCTGCGCGAGTGAAGCCTCGGAGAGAGCAGCTTGTG 1714
 QY 1760 AGAGCAGAGCGGCTCTGTGGGCGAGCAGTCAATTTCAATGAGTAAAGCAGCAAAAGA 1819
 DB 1715 AGAGCAGAGCGGCTCTGTGGGCGAGCAGTCAATTTCAATGAGTAAAGCAGCAAAAGC 1774
 QY 1820 CTG-----CTGCTCCCTCGCGCTCACTGTGTATTAATGTATGATCACTAGGCGG 1870

DB 1775 CAGAGAGAGCAGTCTGCACTCTGCGCAACCCATTTAGAGTAATGTATGATCAACCAAGG 1834
 QY 1871 TTGGCTCTCTGAGAGCCCTCCCTGCTTCTTCTGATAGCCATCCCTGGGCTGTCAAGAGC 1930
 DB 1835 CTGGCTCTCCAGAGGCGCTCCCATCATCT-----GGGTGCTGCTGGGCGGCTGTCTAGAGC 1889
 QY 1931 TGGAGTTGAGCTTTGTGTAACTGATACAGAGCAGGCTGACCATCAGCGGAGAG 1990
 DB 1890 CAG-----AGTCTGGGCTGTCTGTGCTGATGATGAGCAGAGC---TACTGCCAGC 1940
 QY 1991 CAGAGCCATGTCCAGAGAGCTCTCTGCTGCTGTGCTGATCCCTAGTCTGTGAGACT 2050
 DB 1941 CAGAGCAGATACAGAGATGCTCTGAGAGCCAGAGTGTCTGCTAGCCTGTGAGAGCT 2000
 QY 2051 TCTGTGCACTGTTTCCAAAGCTGTAACCTGACTGT 2088
 DB 2001 CTCTGGGTGTCTCCAAAGAGCTGTGAGGCTCATGT 2038

RESULT 10
 BD244713 2867 bp DNA linear PAT 17-JUL-2003
 LOCUS
 DEFINITION Genes encoding for the human and murine death inducer-obli-
 terator-1.
 ACCESSION BD244713
 VERSION BD244713.1 GI:33054483
 KEYWORDS JP 2002526040-A/2.
 SOURCE
 ORGANISM Mus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2867)
 AUTHORS Alonso,C.M., Domingo,D.G., Grandien,A., Leonardo,B. and Martinez,P.
 TITLE Genes encoding for the human and murine death inducer-obli-
 terator-1
 JOURNAL JOURNAL
 COMMENT
 OS Mus sp. (murine)
 PN JP 2002526040-A/2
 PD 20-AUG-2002
 PF 10-SEP-1998 JP 2000570314
 PR 10-SEP-1998 SR 9803069-5,17-SEP-1998 US 60/100873 PI
 CARLOS MARTINEZ ALONSO, DAVID GARCIA DOMINGO, ALF GRANDIEN, PI
 ESTHER LEONARDO.
 PI PEDRO MARTINEZ
 PC C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P3/10,A61P17/00, PC
 A61P29/00,
 PC A61P35/00,A61P37/06,A61P43/00,C07K14/47,C07K16/18,C12N1/21, PC
 C12N5/10,
 PC C12P21/02,C12Q1/02,C12Q1/68,G01N33/15,G01N33/50//C12P21/08, PC
 C12N15/00,
 PC C12N5/00,A61K37/02
 CC Genes encoding for the human and murine death inducer- CC
 FH Key obli-
 terator-1
 FT source location/Qualifiers
 1..2867 /organism="Mus sp. (murine)".
 1..2867 Location/Qualifiers
 1..2867 /organism="Mus sp."
 /mol_type="genomic DNA"
 /db_xref="taxon:10095"

FEATURES
 source
 location/Qualifiers
 1..2867 /organism="Mus sp."
 /mol_type="genomic DNA"
 /db_xref="taxon:10095"

ORIGIN
 Query Match 39.2%; Score 1023.6; DB 6; Length 2867;
 Best Local Similarity 71.9%; Pred. No. 8.7e-264;
 Matches 1494; Conservative 0; Mismatches 534; Indels 50; Gaps 10;

QY 21 TCCGCGCGCTTGGGAGAAATGCGTGGAGACCTTGAAGAGCTGTGAGAGCTTACTCCAGG 80
 DB 1 TCCGCGCGCTTGGGAGAAATGCGTGGAGATCCGCGCGCGGAGAGCTTGTTAAGAGG 60
 QY 81 GAACAGCCTTAGATAATCTGAGTGTGTAAGAAATACGAAGCCTGTACTGTGAAACAGTG 140

REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
ORIGIN

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Leonardo, B.; Martinez, P., Alonso, C.M., Domingo, D.G. and Grandien, A.
Genes encoding for the human and murine death in
ducer-oblierator-1
Patent: WO 0015787-A 2 23-MAR-2000;
LEONARDO ESTHER (ES) ; MARTINEZ PEDRO (ES) ; CONSEJO SUPERIOR
INVESTIGACION (ES) ; DOMINGO DAVID GARCIA (ES) ; MARTINEZ ALONSO
CARLOS (ES) ; BANNERMAN DAVID GARDNER (GB) ; GRANDIEN ALP (SB)
Location/Qualifiers
1. 2867
/organism="Mus sp."
/mol_type="unassigned DNA"
/db_xref="taxon:10095"

Query Match 39.2%; Score 1023.6; DB 6; Length 2867;
Best Local Similarity 71.9%; Pred. No. 8,7e-264;
Matches 1494; Conservative 0; Mismatches 534; Indels 50; Gaps 10;

QY 21 TCCGCGCGGTTCCGGGAAATGGCTGCGAGACCTTAGAGCGCTGCGAGCTTAATCCACGG 80
DB 1 TCCGCGGAGCTCGGAAATGGCTGCGGATCCCGCGCGCGGAGCTGTTTAAAGG 60
QY 81 GAACAGCTCTAGATTAATCTAGATTGAAATACGAAGCTCTTACTCTGAAACAGTG 140
DB 61 CAGTCCCACTGTCCCTTGTGTGTGACCTCCGAAATCTTCTATTGAACTGTG 120
QY 141 GCTGACACAGGTTGTTGTGAGCTGCTGTCTGCTTGAACCCAGAGTTTCGTCC 200
DB 121 ACTGACACAGTGGGTAG--GCTTGGCCGTCTGCTTAC-----TGGCC 166
QY 201 AGGGTTTTGTGTTATTAGGATT--CAGGAAAGTGTCCAGCTTCACTGTGAG 259
DB 167 CAGGCTATAATTATATGATGATTTCCAGCCAAAGTTTCCAGCTTCACTGTGAG 226
QY 260 CAGGTATGAGACAAAGGCCAATGAGGAGTGAAGGAGCTTAAGGCTTAACCA 319
DB 227 CAGGTATGAGATTAAGGAGCCTGAGCAATGAGAGAGCAACCCAGGCTTAACCA 286
QY 320 CCAGCAAAAGATTGAGAAACATGGGGTTTTCGAAGACCACTATCCCAAGCAGAG 379
DB 287 CCAATGAGAGATTGAGAAACCTGGGTTTTCGAAGACCACTATCCCAAGCAGAG 346
QY 380 GCGAGGAGAGCGCGAGGCTGAGCCACTGAGCCGCGCAACCCCAAGAGAGCTGG 439
DB 347 GTGAGAGAGACACGAGAGCGAGCCCACTGAGCAGCAAC-----ACAGCAGTA 400
QY 440 TGTCTCTGCGCGAGTGGAGAGCAGCCCAAGCGCACTGAGCGCTGAGCACTT 499
DB 401 TCTCTCTGCGCGAGTGGAGAGCAGCAACCAACCTGAGAGGAGTGAAGAGTT 460
QY 500 CCATTGCGCGCGCGCGAGAGAGCATGCTGTCTCCCTGAGGATTCTGTGAGC 559
DB 461 CCAAGGTTCCGCGCGCGAGAGAGAGATGTCGCGGTCTCTGAGAGATTCACTGAGC 520
QY 560 CCAGCTCTGCGCGCGCGCAAGACCCAGACAGCTCTGAGGAGCGCTGAAAGCCT 619
DB 521 CCAATCTTCCAGAGTCACTGATGAGAGAGCTTCCAGGAGGAGCTGTAACAGTT 580
QY 620 CTGAGACCAAGAGCGCGCGCGAGTCTTCCAGCTGTGAAGAGAGCAAGCTCT 679
DB 581 CTGAGATCAGAGAGGCTCTGTATCTCTTA--GGAAACCAATCTCTGCTCTT 637
QY 680 CTGAAAGAGTGAAGAGAGATGACACAGTGAATCTCCGATATGATCAAGCTGAGC 739
DB 638 CTGAAAGAGCAAGAGAGATGAGAGAGAGAGACCTCTGAGAGTGAAGTGAAGC 697
QY 740 TGACCTTGAAGAGCTTGAAGATGCTTTCGAGAGAGCGGAGACAGAGCTCACTGAG 799
DB 698 TTAGGTTGAAGAGACTTCAAGACCGCTTCGAGAGAGAGAGAGAGAGAGCTGAGAG 757
QY 800 GGCCCTGAAGAGATCAAGAGTCCCTGCGGAGAGAGCGCGGAGAGGCTCCGCGG 859

DB 758 GGTCTCTAGAGGAGATGAGATTCGCTGAGAGAGAGAGAGAGAGATTCGCGG 817
QY 860 AGACTGTGGCTCCGAGGCGCAGTGAACCTGTGAGAGGCTCTCCGCACTAAGAGAGC 919
DB 818 AATCTGGAGATGTCCTAATTAAGCAGTGGCCGAGCAGAGACAGCTCTCTGTAAGCAGAGC 877
QY 920 CCGAGAGAGATCAAGAGGAGTGTGCTCCAGGCTGGGAAGATGACAGAGAGATGAG 979
DB 878 CTGAGGCTAGTCAAGAGACAGATGCTCCAGTGAAGACAGATGACATGAGAAATCAGTTGG 937
QY 980 AGGAGAGAGCGGCTCAGGACATCAAAAGATGAGAGACCTGAGACTTGGCCGAGCGAAGC 1039
DB 938 AAGGAGAGCGCATCAGAGAAATACAGAGAGAAACCCAGAGAGACGGGCAAAACCAAGC 997
QY 1040 CTGATGTGAGAGGTTACAGACCCCAAGCGCTGTATGATGATTTCCGCAAGCTCAACAC 1099
DB 998 CTGAGGTGAGAGTTTACAGACCCCAAGCTGTATGATGATGCGGCAAGCTCAACAC 1057
QY 1100 ACAGGTTATGATTTGCTGTGACCGCTGTGAAGATGGTTTCAAGGCAATGTTGTGGGCA 1159
DB 1058 ACAGGTTATGATCTGCTGTGATCGTGTGAGAGTGGTTCCATGCTGACTGTGGGTA 1117
QY 1160 TTTCTGAGGCTCAGAGGAGCTTTTGAAAGAAATGGGAAAGACTATATCTCCCAACT 1219
DB 1118 TTTCTGAGGCTCAGAGGAGCGGCTCTGAAAGAGACGGGAGAGACTATCTGCCAAT 1177
QY 1220 GCACCATTTCTGCAAGTGAAGATGAGACTATTCAGAAACCGCAGATCAGCAGAAAGCTA 1279
DB 1178 GCACCATTTTGAAGTGAAGATGAGACCAACCGTAGGCGCCACCAATGAGCAGAGCTCTG 1237
QY 1280 AATGAGACCTGAGATCTGTATGAGACCAATGATGATGATGATGATGATGATGATGATG 1339
DB 1238 GGTGAGATCTGT 1297
QY 1340 AGTCTAGCGAGAGACCAAGAGATTAAGGATGATGATGATGATGATGATGATGATGATG 1399
DB 1298 AGTCCGAGAGAGACCAAGAGATTAAGGATGATGATGATGATGATGATGATGATGATG 1357
QY 1400 AGAAGAACTCAAGATCTTCCAGCTGTGATGAGGCGCTGTGTCTTCAAAATGATG 1459
DB 1358 AGAAGAACTCAAGATCTTCCAGCTGTGATGAGGCGCTGTGTCTTCAAAATGATG 1417
QY 1460 GCGCGGCTGT 1519
DB 1418 GCGCGGCTGT 1477
QY 1520 TCAACACGCGCAGCAGATGAGATTTCTAAGCTCAGGTAAAGAGAGAGAGAGAGAG 1579
DB 1478 TCAACACGCGCAGCAGATGAGATTTCTAAGCTCAGGTAAAGAGAGAGAGAGAGAG 1537
QY 1580 CTAAAGAAAGATGAGATGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1639
DB 1538 CCAAGAGAAAGATGAG 1597
QY 1640 CAGGTATTAATCTCTCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1699
DB 1598 TGGGATTAATCTCTCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1657
QY 1700 TGAAG 1759
DB 1658 TGAAG 1714
QY 1760 AG 1819
DB 1715 AG 1774
QY 1820 CTG-----CTGCTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1870
DB 1775 CAG 1834
QY 1871 TTGGCTCTGAG 1930

Db 1835 CTGGCTTCCAGGCCCTCCCATCATCT-...-GGTGGCTGCTGGGGCTGTCTAGAGC 1889

Qy 1931 TGGAGTTCAGAGCTTTGTGTTAAGCTATCAGACACCGGCTGCACATCAGCGGAG 1990

Db 1890 CAG-----AATCTGGGTGTTCTGTCTGTATATAGCAGCAGC-...-TCACTGCCAGC 1940

Qy 1991 CAGAGCCCATGTCCAGATGCTCTCTGCTGCTGCTGTGTCATCTCTTACTGTCTAGACT 2050

Db 1941 CAGAGCAGATACCAAGATGCTCTGAGACCCAGATGTTCTGCTGCTAGCTGTGAGCCT 2000

Qy 2051 TCTGTACACTGTTTCCAAAGCTGTAAACCTCACTGTGT 2088

Db 2001 CTCTGGGTGTTCTCTAAAGAGCTGTAGGCTCATGT 2038

RESULT 12

AKI29117 4906 bp mRNA linear ROD 21-NOV-2003

LOCUS DEFINITION Mus musculus mRNA for mKIAA0333 protein.

AKI29117

VERSION AKI29117.1 GI:37359897

KEYWORDS FLI CDNA.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1

AUTHORS Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Koseki,H., Hirooka,S., Suga,Y., Nagase,T., Ohara,O. and Koga,H.

TITLE Prediction of the coding sequences of mouse homologues of KIAA gene: III. the complete nucleotide sequences of 500 mouse KIAA-homologous cDNAs identified by screening of terminal sequences of cDNA clones randomly sampled from size-fractionated libraries

JOURNAL DNA Res. 10 (4), 167-180 (2003)

MEDLINE 22972043

PUBMED 14621295

REFERENCE 2 (bases 1 to 4906)

AUTHORS Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.

TITLE Direct Submission

JOURNAL Submitted (23-JUL-2003) Hisashi Koga, Kazusa DNA Research Institute, Laboratory for Genome Informatics; 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918) The CREATE program supported by Japan science and technology corporation; cDNA full insert sequencing; Kazusa DNA Research Institute; cDNA library construction, clone selection and 5'- & 3'-end one pass sequencing.

COMMENT

FEATURES

Source

1..4906

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="mbg16015"

/tissue_type="brain"

/dev_stage="adult"

/note="vector:modified pBC SK+"

1..4906

/gene="mKIAA0333"

<224..3829

/gene="mKIAA0333"

/note="CDS is predicted by in silico analysis. Start codon is not identified."

/codon_start=1

/evidence=not experimental

/product="mKIAA0333 protein"

/protein_id="BAC97927.1"

/db_xref="GI:37359898"

/translaction="FYVGFPAKGFPAFSGTGMDDKGLSNBEAPKAIKPTSEKPRKT WGRRTTIAREGAGDTEVPSBOQPOOHMISLRSRGROPKRTREVEFLTVRRRK KNYVLSBSSSEPTSTVTVETVETBSGVSSEISRGVPSDLSGKEHPSSEAKGK EREEDTSDSDGLTLEKLNRLARRRORRORVSLRSGNRLAKRRERNSATGSGV QISABODRPLCKOEPASGQPPVSGSSTDDIENDEKATGANTGNTENPREAKRPRPC EYDPPNALYCI CRPHNNRMICCRDCEBWFHDCVIGISEAKRLLRNGDYICPNC

gene

CDS

ORIGIN

Query Match 37.7%; Score 985; DB 10; Length 4906;

Best Local Similarity 76.0%; Pred. No. 2,3e-253; Indels 22; Gaps 5;

Matches 1288; Conservative 0; Mismatches 385;

175 GCTTGAACCCAGAGGTTCTGCTGCGACGGTTTGTGTGTAATTAGATT-CAAGGAA 233

187 GCTTGGCGGTCTGCTGTAAGTGGCCCGGCTTAATTTATATAGATTTCCAGCCAA 246

234 AAGTTCAGAGCTTCAAGTGTGAGCAGTATGAGACACAAAGCCACCCAGCATTA 293

247 AGGTTTCCAAAGCTTCAAGTGTGAGCAGTATGAGTATTAAGGCGACCTGAGCAAT 306

294 GGAGGACCTTAAAGCCATCAACCAAGAGCAAGAGTTCAGAAACATGGGGTTTGG 353

307 GGAAGCACCAGAGGTATCAAAACCAAGTAAAGAGTTCAGAAACATGGGGTTTGG 366

354 AAGGACACTATCCGCAAGCAGAGGCGCGAGGGAGCCGAGAGCTGACCCACTGAGCC 413

367 AAGAACCAAGTATGCAACCAAGAGGTGTCAGAGACACAGAGGTGACCCAGTGAAGA 426

414 GCCACCCCAAGCAGAGCAGTGGGCTTCTCTCGCGCGCAGTGGAGGACCCAGAGCG 473

427 GCAAC-----ACAGCAGCATTAACCTCTCCCTCGCGCGCAGTGAACCGCAACAAACG 480

474 CACTGAGCGCTGAGAGCATTCCTGACCATTTGCGGGCGCGCGCAGAGGAGCATGCC 533

481 TACTGAGAGGTGAGAGGTTTCTTACACAGGTTGCGCGCGAGGAAAGAAATGTC 540

534 TGTCTCCCTGAGAGATTGTGAGAGCCACAGTCTGCCCCCGCAGACAGCGCGAGACG 593

541 GGTGTCTCTGAGAGATTCAAGTACCCACATCTTCCACAGTCACTGATGTGAGACAGC 600

594 CTCCAGGCGAGCGCTGAGAGCGCTTCTGAGACAGAAAGCGGCCCCAGTCTGCTTCCAC 653

601 TTCCAGAGGAGCGCTGAGAAACAGTTCTGAGATCAGAAAGTGCCTGTATCTGACTCTT 660

654 AGCTGTGAAGAGACAGCGCTCTTCTGAAAGGTGAAGAGGGGAGTACCAAGATGA 713

661 A---GGAAAGAACATCTGCTCTTCTGAAAGGCAAAAGAGGTGAAGAGAAAGAAAGA 717

714 CACCTCCATAGTGAACAGCATGAGCTGACCTTGAAGAGCTTCAGAAATCGCTTGCAG 773

718 CACTCTGACAGTGAACAGTATGAGCTTACGTTGAAGGAACTTCAGAACCGGCTTCGGAG 777

774 GAAGCGGAGACAGAGCCCATGAGAGGCGCTTGAAGAGGATTCAGAGTCCCTTCGGAA 833

778 AAAGCAGAGAGCAAAACCTGTGAGAGAGTCCCTGAGAGCGAGTCAAGAAATCGCTTGAAGAA 837

834 GAAGCGCGGAGAGAGGTCCTCCGCGAGAGCTGTGGCGCTCCAGAGCGCAAGTACACTGTGGA 893

838 GAAGCGCAGAGAGAGAGTTTCTGCGAAATCGGAGAGTCTCAAAATAGCAAGTGCAGAGCA 897

894 GGGGCTCTGCGCAGTAAAGCAGAGCCCGAGAGAGATCAGAGGAGGTTGTGTCCCAAGCTGG 953

898 GAGCAGACCTCTCTGTAAAGCAGAGACCTGAGGCTAGTACAGGAGCAAGTGTCCCAATCAGA 957

TTTAVDPTNGASATBEDSGRSGVAGDSTGCTGTGVRKSGEDGIGRIERKANPS
GKKKTKTFQPVYALPAPKCTIGPCGSSVAPGDPSTYCSNDCLIKHAAATPRFLSQRGK
KTKRKERKTKPERKSLPKCSQVGIKLSVHKILASKEKREKREKREKREKREKREK
EACBESSTPWSADHNTYAVBKEKREKREKREKREKREKREKREKREKREKREKREK
KALPSALVAGCTSPRLVYKPKLPPYSMAAKAPAKIKLPKSGFKGTIKPRPWPASVTL
GTSRAGSIVGPTWPTAPSKLPGSAAVGTGTPKMSANVPAAPGLPVPVSPAPQPN
SOIKONRIRSLKILMKRYVNDLIMRENVGTLALHIEKEMFLPVOTDRVYSKY
RSIMPNLDPKNOGLFHEVYVBERISLAKYPMKRELYSKELSWTTEKPTSVISRT
KLINESKKNTPKEPTIPDMEDSPVSDSEBEPAPKXKLASSKEDPKREHDSPPVAVNTA
SOHRALFDLNCXICTGVPSBDEPAPKXKLASSKEDPKREHDSPPVAVNTA
DEGLADTLPERNASDPDESTSLNOERKCFPESPDSDHPEPSLGLTSPSSASGSGV
VTTVMGSRPRTALSQSCVTYASWALHDSOSSETLDMIKPALTVASVYSKILAK
PSSPDPRLYSVPSPSPSISERSPREGDTTLPSRLNTIMGTIMOGVAKREYAV
VSGCLDUSBDLPPTIHIGRIAPRTVVDYTGKLSYSKELCLRRFPATBEBAY
ISLVSYSRSRGRFVANNRVRDLYLPLSADPPVSKLPFGGPHVPSGR


```

QY 954 GAAAGATGACGAGAGAGTAAGTTGGAGGAGGAGGCGCTCAGACATCAAGATGAGGA 1013
DB 958 GACAGAGATGACATGAAAAATCAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1017
QY 1014 GCCTGAGACTTGGGCGGAGCCGAGAGCCGATGATGAGGGGTACACACCCCAAGCCCTGTA 1073
DB 1018 CCCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1077
QY 1074 TTGCATTTGCGCGAGGCTTCAGACAAACAGTTTATGATTTGCTGAGACCGCTGTGAGA 1133
DB 1078 CTGCATCTGCGCGAGGCTTCAGACAAACAGTTTATGATTTGCTGAGTGGGTGTGAGGA 1137
QY 1134 ATGTTTATGAGGAGTGTGTGGGCAATTTGAGAGCTCGAGGAGGCTTTTGGAGGAGA 1193
DB 1138 GTGTTTATGAGTGTGTGGGCAATTTGAGAGCTCGAGGAGGCTTTTGGAGGAGA 1197
QY 1194 TGGGAGAGACTATATCTGAGCCAACTGACCATCTGCAAGTGAAGATGAGATCTATTC 1253
DB 1198 CCGGAGAGACTATATCTGAGCCAACTGACCATCTGCAAGTGAAGATGAGATCTATTC 1257
QY 1254 AGAAACGAGATCAGAGAGAGAGGCTTAAATGAGAGCTGAGAGATCTGATGAGCAGATTG 1313
DB 1258 TAGGCGACAGGATGAGAGAGAGGCTTAAATGAGAGCTGAGAGATCTGATGAGCAGATTG 1317
QY 1314 TACAGATATGAGAGAGATGAGAGAGAGGCTTAAATGAGAGCTGAGAGATCTGATGAG 1373
DB 1318 CACAGATATGAGAGAGATGAGAGAGAGGCTTAAATGAGAGCTGAGAGATCTGATGAG 1377
QY 1374 TGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1433
DB 1378 TGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
QY 1434 GGGCGCTGAGTCTCAAAATGATTTGGCGCGGCTGTGACAGTGGGCGAGCCGAGCTC 1493
DB 1438 GGGCTGAGTCTCTTAAATGATTTGGCGCGGCTGTGACAGTGGGCGAGCCGAGCTC 1497
QY 1494 GGTCTATGAGAGATGATCTTAAATGATTTGGCGCGGCTGTGACAGTGGGCGAGCCGAG 1553
DB 1498 TGTGATTTGAGATGATCTTAAATGATTTGGCGCGGCTGTGACAGTGGGCGAGCCGAG 1557
QY 1554 CTCAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1613
DB 1558 TTCAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
QY 1614 CAGTCTTCCGAGATCGGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1673
DB 1618 CAGTCTTCCGAGATCGGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677
QY 1674 AGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1733
DB 1678 AGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1734
QY 1734 AGCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1793
DB 1738 GACTCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1794
QY 1794 TTACATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1844
DB 1798 CTACATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1844
QY 1845 GTATTAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1854
DB 1855 GATTAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1854

```

RESULT 13
 LOCUS HS885L7/c 160241 bp DNA linear PRI 19-APR-2001
 DEFINITION Human DNA sequence from clone RPS-885L7 on chromosome 20q13.2-13.33
 Contains ESTs, STSs, GSSs and eight Cpg islands. Contains the 3' end of the NTSR1 gene for high affinity neurotensin receptor 1, a putative novel gene, a novel gene similar to a fly gene, the gene for opioid growth factor receptor (7-60 protein), the COL9A3 gene

for collagen IX alpha 3, a putative novel gene similar to a fly gene, the TGF5L gene for basic helix-loop-helix transcription factor-like 5, an ARF4 (ADP-ribosylation factor 4) pseudogene, a novel gene and the 3' end of the DATF1 gene encoding death associated transcription factor 1, complete sequence.
 AL035669
 AL035669.43 GI:8979786
 HNG: 7-60: ARF4; COL9A3; collagen; Cpg island; DIO-1; KIAA0333; neurotensin; NTSR1; opioid growth factor receptor; TGF5L.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 160241)
 Smith, M.
 Direct Submission
 Submitted (19-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Jul 8, 2000 this sequence version replaced gi:8919619.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Mp, MORNBP; Information on the MORNBP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>
 This sequence is the entire insert of clone RPS-885L7. The true left end of clone RPS-885L7 is at 140192 in this sequence. The right end of clone RPS-885L7 is at 44512 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RPS-885L7 is from the library RPS1-5 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pCYPAC2.

FEATURES

source

1. 160241
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="20"
 /map="q13.2-13.33"
 /clone="RPS-885L7"
 /clone_id="RPS1-5"
 2. 64

repeat_region

65..353

repeat_region

354..397

repeat_region

3862..11948

gene

3862..11948

gene

4063..7443..7533..9195..11948

gene

4063..7443..7533..9195..11948

gene

4063..7443..7533..9195..11948

gene

4063..7443..7533..9195..11948

gene

4063..7443..7533..9195..11948

```

CDS
/evidence=not_experimental
join(3862..4063,7443..7533,9195..9444)
/gene="MYSR1"
/note="Continues in BA93B14 (Em:ALJ57033) as BA93B14.3
match: proteins: Tr:088319 Sw:P30989 Sw:P20789"
/codon_start=1
/evidence=not_experimental
/product="cd885L7.1.1 (neurensin receptor 1 (high
affinity), isoform 1)"
/protein_id="CAC12747.1"
/db_xref="GI:12313997"
/db_xref="GCA:P30989"
/db_xref="SWISS-PROT:P30989"
/translatability="NTMSPFIFPMVAVSVLNTIIANKLTVMVRQAAEQGVCTGGE
HSTFMALEPGRVQALRHGVRVLRALVAVLCMLPHYRRLAMCYISDSQMPFLYD
PHYFMVMTNALPYVSTTNPLTLYLVNANFRHIFLATIACLCVPMRRRRRPAFRK
ADVSNNHTLSNATRETTY"
4198..5157
/note="12 copies 80 mer 61% conserved"
/misc_feature
complement(4223..5153)
/note="Random repeat. region contains two forced joins and
single clone. Assembly consistent with restriction
digest."
4229..5114
/note="443 copies 2 mer cc 59% conserved"
4231..5166
/note="2234 copies 4 mer tccc 60% conserved"
4240..5151
/note="19 copies 48 mer 61% conserved"
4248..5159
/note="38 copies 24 mer 61% conserved"
4264..5163
/note="15 copies 60 mer 61% conserved"
4286..5152
/note="17 copies 51 mer 59% conserved"
4308..5171
/note="27 copies 32 mer 61% conserved"
4317..5156
/note="21 copies 40 mer 62% conserved"
4344..5142
/note="17 copies 47 mer 59% conserved"
4348..5145
/note="42 copies 19 mer 60% conserved"
4431..5144
/note="21 copies 34 mer 60% conserved"
4456..5148
/note="21 copies 33 mer 60% conserved"
4486..5151
/note="9 copies 74 mer 69% conserved"
4530..5087
/note="18 copies 31 mer 71% conserved"
4533..4822
/note="5 copies 58 mer 77% conserved"
4720..5137
/note="11 copies 38 mer 61% conserved"
4833..5064
/note="4 copies 58 mer 79% conserved"
5051..5164
/note="19 copies 6 mer tctctc 63% conserved"
join(5808..5912,7443..7533,9195..11948)
/gene="MYSR1"
/product="cd885L7.1.2 (neurensin receptor 1 (high
affinity), isoform 2)"
/note="match: ESTs: Em:ALJ570553"
/evidence=not_experimental
5934..5864
/note="11M4 repeat: matches 4509..5488 of consensus"
6974..7049
/note="4 copies 19 mer 80% conserved"
7224..7451
/note="6 copies 38 mer 98% conserved"
7892..8074
/note="MER68A repeat matches 68..210 of consensus"

```


OY 1426 ----- 1425
 DB 481 CCAGCTCCTGCTCTCTGAGGATGGAATTGCTGTCTGAAGACATCTCTGCT 540
 OY 1426 ----- GTGATAGAGGCGCTGCTGCTCAAAATGTA 1456
 DB 541 TCACACTCTTGACATGACATTAAGCTGCAAGCTGATAGAGCGCTGCTGCTCAAAATGTA 600
 OY 1457 TTGGCCCCGGGTGCTGTCACTGTGGCGAGCCGACTCGGTGACTGCACTAATGACTGTA 1516
 DB 601 TTGGCCCCGGGTGCTGTCACTGTGGCGAGCCGACTCGGTGACTGCACTAATGACTGTA 660
 OY 1517 TCTTCAAAACAGCCGACGACGACATGAACTTCTAAGCTCAGGTAAAGAACAGACC 1576
 DB 661 TCTTCAAAACAGCCGACGACGACATGAACTTCTAAGCTCAGGTAAAGAACAGACC 720
 OY 1577 AGCCTAAAGAAAAGTAAAGTAAAGGACGAGAACCCGACTCTTCCGAAATGCGGTGCTC 1636
 DB 721 AGCCTAAAGAAAAGTAAAGTAAAGGACGAGAACCCGACTCTTCCGAAATGCGGTGCTC 780
 OY 1637 AGGCAAGATTTAAATCTCTTCTGTGCAACAGACACAGCTCCAGAAAAAAGAGACCA 1696
 DB 781 AGGCAAGATTTAAATCTCTTCTGTGCAACAGACACAGCTCCAGAAAAAAGAGACCA 840
 OY 1697 CAGTAAAGAGGAGAGTGTGTCTCTGCGGAGTGAAGACACTCGGAAAGAACAGCTT 1756
 DB 841 CAGTAAAGAGGAGAGTGTGTCTCTGCGGAGTGAAGACACTCGGAAAGAACAGCTT 900
 OY 1757 GTGAAGAGACGACCGCTGTGTGGGAGAGGATCAAAATTAAGTAAAGCCAGAA 1816
 DB 901 GTGAAGAGACGACCGCTGTGTGGGAGAGGATCAAAATTAAGTAAAGCCAGAA 960
 OY 1817 AGACTGCTGCTCTCTGCGCTGACATCTTGTATTAATCCACGAA 1859
 DB 961 AGACTGCTGCTCTCTGCGCTGACATCTTGTATTAATCCACGAA 1003
 RESULT 15
 LOCUS HSJ563B14/c 28519 bp DNA linear PRI 19-APR-2001
 DEFINITION Human DNA sequence from clone RP4-563B14 on chromosome 20. Contains the 5' of the DATP1 gene encoding the death associated transcription factor 1, the 5' end of a novel gene, ESTs, STSs, GSSs and four Cpg islands, complete sequence.
 ACCESSION AL117379
 VERSION AL117379.14 GI:7329921
 KEYWORDS HTG; Cpg island; DATP1.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 28519)
 AUTHORS Collier, R.
 TITLE Direct SubMISSION
 JOURNAL Submitted (19-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1BA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 COMMENT On Mar 26, 2000 this sequence version replaced gi:7327990. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPP; Information on the WORMPP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpp. This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/BGP/Chr20
 IMPORTANT: This sequence is not the entire insert of clone RP4-563B14. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-305P22 is at 28420 in this sequence. The true right end of clone RP5-885L7 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP4-563B14 is from the library RP4-4 constructed by the group of Pletter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pCYPAC2.

FEATURES

source
 Location/Qualifiers
 1..28519
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="20"
 /clone="RP4-563B14"
 /clone_1bp="RP4-4"
 complement(1)..26931)
 /gene="DATP1"
 /product="d5563B14.1 (Death associated transcription factor 1 (contains KIAA0333) isoform 1)"
 /note="match: cDNAs: Em: AK002127 Em: AJ238332
 match: ESTs: Em: AW410673 Em: AA050179 Em: AA43704
 Em: N55493"
 /evidence="not experimental"
 complement(1)..648)
 /gene="DATP1"
 /note="continues in clone d5885L7 (Em: AL035669) as gene d5885L7.9
 match: proteins: Tr: Q9WV00"
 /codon_start=1
 /evidence="not experimental"
 /product="d5563B14.1 (Death associated transcription factor 1 (contains KIAA0333) isoform 1)"
 /protein_id="CA890275.1"
 /db_xref="GI:7711029"
 /db_xref="GOA:Q9BTC0"
 /db_xref="SWISS-PROT:Q9BTC0"
 /translation="MDKQDPSENEAPKAIKPSKEPRKTRGRTTIARBEAGDAEADLPPEPPQQLGLSLRSGRQPKTERVEBPLTIARRGRKMPVSLSDSEPTSCPATDARTSSEGVSESTRSGPOSASTVVKRPPASBRVKGDDHDYSDSDIGITLKEILNRLARRKKEBEPTEPLKGIQSLRLKRRREBGPATVGSASDITVEGVLPSTK39..561
 /note="Cpg island"
 /evidence="not experimental"
 complement(1)..426..650,3246..3442,15475..>15507)
 /gene="DATP1"
 /product="d5563B14.1 (Death associated transcription factor 1 (contains KIAA0333) isoform 1)"
 /note="Death associated transcription factor 1 (translation of cDNA PLACE1009158 (Em: AK002127)) isoform 2 alternatively spliced in 5' UTR
 match: ESTs: Em: AW410673"
 /evidence="not experimental"
 complement(1177..1649)
 /gene="DATP1"
 /note="match: GSS: Em: AQ127293"
 1327..1417
 /note="MULTI repeat: matches 439..535 of consensus"
 1462..1514
 /note="AluSc repeat: matches 246..298 of consensus"
 1749..2041
 /note="match: GSS: Em: AQ253731"

repeat_region	1786. .2093	/note="Aluux repeat: matches 1. .313 of consensus"
repeat_region	2898. .2953	/note="MER2 repeat: matches 97. .153 of consensus"
misc_feature		/note="MER2 repeat: matches 97. .153 of consensus"
repeat_region		/note="MER2 repeat: matches 97. .153 of consensus"
repeat_region	4231. .4366	/note="MER2 repeat: matches 97. .153 of consensus"
repeat_region	4520. .4828	/note="Aluud repeat: matches 3. .145 of consensus"
repeat_region	5459. .6077	/note="Aluud repeat: matches 1. .302 of consensus"
repeat_region	6078. .6366	/note="Charliel repeat: matches 2193. .2759 of consensus"
repeat_region	6367. .6994	/note="Aluub repeat: matches 1. .266 of consensus"
repeat_region	7348. .7470	/note="Charliel repeat: matches 4. .2193 of consensus"
repeat_region	8302. .8594	/note="Aluub repeat: matches 136. .259 of consensus"
repeat_region	9647. .9932	/note="Aluub repeat: matches 1. .301 of consensus"
repeat_region	9933. .10231	/note="Charliel repeat: matches 2. .239 of consensus"
repeat_region	10232. .10392	/note="Aluug repeat: matches 1. .300 of consensus"
repeat_region	10669. .10836	/note="Charliel repeat: matches 239. .523 of consensus"
misc_feature	10781. .10901	/note="Aluub repeat: matches 135. .301 of consensus"
misc_feature	10898. .11020	/note="match: STS: Em:G32962"
repeat_region	10956. .11135	/note="match: GSS: Em:AQ387275"
repeat_region	11138. .11305	/note="3 copies 60 mer 78% conserved"
repeat_region	11146. .11309	/note="14 copies 12 mer 67% conserved"
repeat_region	11146. .11309	/note="14 copies 12 mer 67% conserved"
misc_feature		/note="41 copies 4 mer atac 64% conserved"
repeat_region		complement(11948. .12516)
repeat_region		/gene="DATP1"
repeat_region		/note="match: GSS: Em:AQ415373"
repeat_region	12058. .12208	/note="match: GSS: Em:AQ415373"
repeat_region	12213. .12509	/note="PRAM repeat: matches 1. .144 of consensus"
misc_feature	12555. .13045	/note="Aluux repeat: matches 1. .301 of consensus"
misc_feature	13880. .14299	/note="match: GSS: Em:AQ804060"
repeat_region	14246. .14317	/note="match: GSS: Em:AQ132474"
misc_feature	14712. .15898	/note="L2 repeat: matches 2581. .2658 of consensus"
misc_feature		/note="CPG island"
misc_feature		/evidence=not experimental
misc_feature		complement(15147. .15331)
repeat_region		/gene="DATP1"
repeat_region		/note="match: GSS: Em:AP190856"
repeat_region	15283. .15310	/note="match: GSS: Em:AP190856"
repeat_region	15515. .15636	/note="7 copies 4 mer cccg 96% conserved"
misc_feature	18118. .18638	/note="2 copies 61 mer 86% conserved"
misc_feature		/note="CPG island"
misc_feature		/evidence=not experimental
misc_feature		complement(18567. .19431)
repeat_region		/gene="DATP1"
repeat_region		/note="match: GSS: Em:AQ637346"
repeat_region	18989. .19268	/note="match: GSS: Em:AQ637346"
misc_feature		/note="Aluux repeat: matches 2. .312 of consensus"
misc_feature		complement(18996. .19482)

Query Match	25.1%	Score 654.6	DB 9	Length 28519
Best Match Similarity	98.7%	Pred. NO. 2.7e-164		
Matches 660	Conservative	0	Mismatches 9	Indels 0
			Gaps	0

Query Match	25.1%	Score 554.6	DB: 9	Length 28519
Best Local Similarity	98.7%	Pred. No. 2.7e-164		
Matches 660:	Conservative	0	Mismatches 9	Indels 0
				Gaps 0
QY	244	GCTTTCAGTGTGGAGCAGGTAATGAGACGACAAAGGCGACCCGACCAATGAGAGGACCT	303	
DB	669	GCTTTCACCTTTACCGCTAGGTAATGAGACAAAGGCGACCCGACCAATGAGAGGACCT	610	
QY	304	AAGGCCATCAAAACCACCAAGCAAAAGATTGAGAAAATATGGGCTTTTCGAGAGCACT	363	
DB	609	AAGGCCATCAAAACCACCAAGCAAAAGATTGAGAAAATATGGGCTTTTCGAGAGCACT	550	
QY	364	ATCCGCAAAGCCAGAGGGCGCAGGGGAGCCGAAAGCTGACCACTGAGGCGGCACTCCCA	423	
DB	549	ATCCGCAAAGCCAGAGGGCGCAGGGGAGCCGAAAGCTGACCACTGAGGCGGCACTCCCA	490	
QY	424	CAGCAGCAGCTGGGACCTGTCCCTCGCGCGCAGTGGAGGACAGCCCAAGCGCACTGAGCGC	483	
DB	489	CAGCAGCAGCTGGGACCTGTCCCTCGCGCGCAGTGGAGGACAGCCCAAGCGCACTGAGCGC	430	
QY	484	GTGAGCAGTTTCTGACCATTTGCGGGCGCGCGCAGAGAGACATGCTGTCTCTCTG	543	
DB	429	GTGAGCAGTTTCTGACCATTTGCGGGCGCGCGCAGAGAGACATGCTGTCTCTCTG	370	
QY	544	GAGGATTTCTGTGAGACCCCAAGTCTTGCCTCCGCGCAACAAGCCGAGACAGCCTCCGAGGGC	603	
DB	369	GAGGATTTCTGTGAGACCCCAAGTCTTGCCTCCGCGCAACAAGCCGAGACAGCCTCCGAGGGC	310	
QY	604	AGCGTGAAGAACGCTTCTGAGACACAGAAAGCGGCCCAAGTCTGCTTCCACAGCTGTGAAG	663	
DB	309	AGCGTGAAGAACGCTTCTGAGACACAGAAAGCGGCCCAAGTCTGCTTCCACAGCTGTGAAG	250	
QY	664	GAACGACCAAGCTCTTCTGAAAAGGTAAAGAGAGGAGATGACACAGATGACACTCCGAT	723	
DB	249	GAACGACCAAGCTCTTCTGAAAAGGTAAAGAGAGGAGATGACACAGATGACACTCCGAT	190	
QY	724	AGTACACAGCAATGGCCCTGACCTTTGAAAAGACTTTAGAAATCCCTTCGAGAGAAACGGGAA	783	
DB	189	AGTACACAGCAATGGCCCTGACCTTTGAAAAGACTTTAGAAATCCCTTCGAGAGAAACGGGAA	130	
QY	784	CAGAGACCACTGAGAGGCCCCCTGAAAAGGATCCAGAGTCCGCTTCGAGAAAGACGCGG	843	
DB	129	CAGAGACCACTGAGAGGCCCCCTGAAAAGGATCCAGAGTCCGCTTCGAGAAAGACGCGG	70	
QY	844	GAGGAGGATCCCGCCAGACTGTGGGCTCCGAGGCTCAGTGCACCTGTGAGGGGCTCTTG	903	
DB	69	GAGGAGGATCCCGCCAGACTGTGGGCTCCGAGGCTCAGTGCACCTGTGAGGGGCTCTTG	10	
QY	904	CCCACTAAG 912		
DB	9	CCCACTAAG 1		

. Tue Apr 27 10:09:37 2004

us-09-787-016a-1.rge

Page 25

Search completed: April 23, 2004, 14:16:14
Job time : 9971.23 secs

CC Inducer-oblierator 1 (DIO-1) polypeptides. The polypeptides can be
CC expressed by standard recombinant methodology. The DIO-1 polypeptides,
CC agonists and antagonists are used as a medicament for treating diseases
CC characterized by an alteration in cell death or by hyperproliferation,
CC e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
CC tumours, malignant tumours or hyperproliferative skin disorders. They are
CC also useful in the treatment of metabolic, proliferative or inflammatory
CC conditions. The present sequence represents a cDNA encoding the murine
CC DIO-1 polypeptide
CC
XX

SQ Sequence 2867 BP; 763 A; 666 C; 774 G; 664 T; 0 U; 0 Other;

Query Match 100.0%; Score 2867; DB 3; Length 2867;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2867; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TCCGTGTAGCTCTGAAATGAGCTCCGGGATCCGGGCGCGGGAGCTTTTAAAGG 60
DB 1 TCCGTGTAGCTCTGAAATGAGCTCCGGGATCCGGGCGCGGGAGCTTTTAAAGG 60
QY 61 CAGTCCCACTGTCTCTTGTGTTTGAAGCTCCGAATCTTCTCATTTAGAACTGTG 120
DB 61 CAGTCCCACTGTCTCTTGTGTTTGAAGCTCCGAATCTTCTCATTTAGAACTGTG 120
QY 121 ACTGACAAAGTGGGGTGAAGGCTTGGCGTCTGAAGCTGAGCCAGGTCTATATTT 180
DB 121 ACTGACAAAGTGGGGTGAAGGCTTGGCGTCTGAAGCTGAGCCAGGTCTATATTT 180
QY 181 TATGTAGATTTCAGCCAAAGTTTCCAGCTTTCAGTTTGGGACAGATATGATAT 240
DB 181 TATGTAGATTTCAGCCAAAGTTTCCAGCTTTCAGTTTGGGACAGATATGATAT 240
QY 241 AAAGGGCACTGAGACATGAGAGACACCAAGGCTATCAACCAACGATAGAGATT 300
DB 241 AAAGGGCACTGAGACATGAGAGACACCAAGGCTATCAACCAACGATAGAGATT 300
QY 301 AGGAAACCTGGGGTTTGAAGAACAGATTTGCAACGATGAGGGTGCAGAGACAG 360
DB 301 AGGAAACCTGGGGTTTGAAGAACAGATTTGCAACGATGAGGGTGCAGAGACAG 360
QY 361 GAGGCGGACCCAGTGAAGACCAACAGACATTAACCTTCCCTGCGCGAGTGA 420
DB 361 GAGGCGGACCCAGTGAAGACCAACAGACATTAACCTTCCCTGCGCGAGTGA 420
QY 421 CGGCAACCAAAAGCTGAGAGGCTAGAGAGTTTCTTACACGCTTCCGCGGAGG 480
DB 421 CGGCAACCAAAAGCTGAGAGGCTAGAGAGTTTCTTACACGCTTCCGCGGAGG 480
QY 481 AAAAAAGATGCGGGTCTCTGAGAGATTCAGTGAAGCCACATCTTCCACAGTCA 540
DB 481 AAAAAAGATGCGGGTCTCTGAGAGATTCAGTGAAGCCACATCTTCCACAGTCA 540
QY 541 GATGTGAGACAGCTTCCGAGGGAGCGTTGAAGCAGTTTCAATGAGAGTGGCT 600
DB 541 GATGTGAGACAGCTTCCGAGGGAGCGTTGAAGCAGTTTCAATGAGAGTGGCT 600
QY 601 GATGTGAGACAGCTTCCGAGGGAGCGTTGAAGCAGTTTCAATGAGAGTGGCT 660
DB 601 GATGTGAGACAGCTTCCGAGGGAGCGTTGAAGCAGTTTCAATGAGAGTGGCT 660
QY 661 GAGGAGAAAGACACTGACAGTGAAGTGAAGCTTGAAGGAACTTCAGAAC 720
DB 661 GAGGAGAAAGACACTGACAGTGAAGTGAAGCTTGAAGGAACTTCAGAAC 720
QY 721 CGCCTTCGGAAGAAAGCGAGAGCAAACTGTGAGAGGTCCTGAGAGGCACTG 780
DB 721 CGCCTTCGGAAGAAAGCGAGAGCAAACTGTGAGAGGTCCTGAGAGGCACTG 780
QY 781 CGCCTTCGGAAGAAAGCGAGAGCAAACTGTGAGAGGTCCTGAGAGGCACTG 840
DB 781 CGCCTTCGGAAGAAAGCGAGAGCAAACTGTGAGAGGTCCTGAGAGGCACTG 840
QY 841 AGTCCGAGAGCAAGACCTCTGTAAAGAGAGCTGAGGCTAGTCAAGGACCAAGT 900
```

```
DB 841 AGTCCGAGAGCAAGACCTCTGTAAAGAGAGGCTGAGGCTAGTCAAGGACCAAGT 900
QY 901 TCCAGTCAAGACAGATGACATGAAATCATGTTGAAGGAGGCGACTCAAGGAAAT 960
DB 901 TCCAGTCAAGACAGATGACATGAAATCATGTTGAAGGAGGCGACTCAAGGAAAT 960
QY 961 ACAGAGAAACCCAGGGAAGCGGGCAACCAAGCTGAGTGAAGTTTACAGACCC 1020
DB 961 ACAGAGAAACCCAGGGAAGCGGGCAACCAAGCTGAGTGAAGTTTACAGACCC 1020
QY 1021 AATGCCCTGTACTGATCTGCTCCGAGCTTCAACCAAGGTTATGATCTGCTGAT 1080
DB 1021 AATGCCCTGTACTGATCTGCTCCGAGCTTCAACCAAGGTTATGATCTGCTGAT 1080
QY 1081 CGGTGTGAGAGTGTCTTCAAGTGAAGTGTGAGTATTTCTGAGGCGCGAGGCTC 1140
DB 1081 CGGTGTGAGAGTGTCTTCAAGTGAAGTGTGAGTATTTCTGAGGCGCGAGGCTC 1140
QY 1141 CTGGAAAGGAACGGGGAAGACATGATGACCAATTTGACCAATTTGCAAGTGAAG 1200
DB 1141 CTGGAAAGGAACGGGGAAGACATGATGACCAATTTGACCAATTTGCAAGTGAAG 1200
QY 1201 GAGACAAACGGTGAAGCCCAATGAGAGGACTCTGAGTGAAGTGTGAGTCTGAT 1260
DB 1201 GAGACAAACGGTGAAGCCCAATGAGAGGACTCTGAGTGAAGTGTGAGTCTGAT 1260
QY 1261 GGCACAGCTGCAACAGATGAGGACATGAGGAGAAAGTCCGAGGAAGACCAAGGCA 1320
DB 1261 GGCACAGCTGCAACAGATGAGGACATGAGGAGAAAGTCCGAGGAAGACCAAGGCA 1320
QY 1321 AAGGTTAGATTGAGAAAGGAGCAAAACCCAGCGCAAGAAATCAAGATATTCAG 1380
DB 1321 AAGGTTAGATTGAGAAAGGAGCAAAACCCAGCGCAAGAAATCAAGATATTCAG 1380
QY 1381 CCTGTGTAGAGGCTCTGAGTCTCTTAATGATTTGAGCTTCCAGTGAAGCA 1440
DB 1381 CCTGTGTAGAGGCTCTGAGTCTCTTAATGATTTGAGCTTCCAGTGAAGCA 1440
QY 1441 CAGCTGAGCTCTGTATATTTGAGATGATGATGATTTCTCAACAGCAGAGCTTCA 1500
DB 1441 CAGCTGAGCTCTGTATATTTGAGATGATGATGATTTCTCAACAGCAGAGCTTCA 1500
QY 1501 AGATTCTAAGTTCAAGTGAAGAAACAAACCAAGGAAAGGTCAAGAGCAAG 1560
DB 1501 AGATTCTAAGTTCAAGTGAAGAAACAAACCAAGGAAAGGTCAAGAGCAAG 1560
QY 1561 CCAAGAAAGTTCAAGTCTTCAAAATGAGTGTTCAGTGGGATTTAAATCTTCTG 1620
DB 1561 CCAAGAAAGTTCAAGTCTTCAAAATGAGTGTTCAGTGGGATTTAAATCTTCTG 1620
QY 1621 CACAGAGACATGAGCTGAGAGAAAGGAAACCCAGTGAAGTGTGCTGCTTCC 1680
DB 1621 CACAGAGACATGAGCTGAGAGAAAGGAAACCCAGTGAAGTGTGCTGCTTCC 1680
QY 1681 AGAGTGAAGCTTCTGAGAGAGGAGCAGCTGTGAGAGCAGACACATCTGAGCAAG 1740
DB 1681 AGAGTGAAGCTTCTGAGAGAGGAGCAGCTGTGAGAGCAGACACATCTGAGCAAG 1740
QY 1741 GACCAACATCAATGCTGTGAAGCAGAGAAAGCCAGAGAGCCCACTGAGCTC 1800
DB 1741 GACCAACATCAATGCTGTGAAGCAGAGAAAGCCAGAGAGCCCACTGAGCTC 1800
QY 1801 ACCCTATTGAGTAAATGTAAGTAAACCAAAAGGCTGCTTCCAGGCGCCCTCCAT 1860
DB 1801 ACCCTATTGAGTAAATGTAAGTAAACCAAAAGGCTGCTTCCAGGCGCCCTCCAT 1860
QY 1861 CTGGGTGCTGCTGAGGCTGTCTAGACACAGAGTCTGAGTGTCTGAGTGAATGA 1920
DB 1861 CTGGGTGCTGCTGAGGCTGTCTAGACACAGAGTCTGAGTGTCTGAGTGAATGA 1920
QY 1921 GCGAGCAGCTCACTGACAGGAGAGAGATCAAGATGCTTGAAGCCAGGCTTTC 1980
DB 1921 GCGAGCAGCTCACTGACAGGAGAGAGATCAAGATGCTTGAAGCCAGGCTTTC 1980
```

```

Db      1921 GCCAGAGCTCATGCGCAGCCAGAAAGACAGATACCAAGATGCTCTGAGACCCAGGTGTC 1980
Qy      1981 CTGGCTAGCCCTGAGAGCCTCTCTGGGTGTTCTTAAAGAGCTGTAGGCTCATGTG 2040
Db      1981 CTGGCTAGCCCTGAGAGCCTCTCTGGGTGTTCTTAAAGAGCTGTAGGCTCATGTG 2040
Qy      2041 GAGGCAATTTCTTATTTGAGTTTGGCCCTTGTGATGATTTGAACTGCTCTCAACAGA 2100
Db      2041 GAGGCAATTTCTTATTTGAGTTTGGCCCTTGTGATGATTTGAACTGCTCTCAACAGA 2100
Qy      2101 TGGGCAACATCAGTGCCTGTTGAGACTCAGCTTGAAGTTCCGACAGGACAGC 2160
Db      2101 TGGGCAACATCAGTGCCTGTTGAGACTCAGCTTGAAGTTCCGACAGGACAGC 2160
Qy      2161 CCAGACAAACCATGAGATTTTCCACGTTACCTATTTTCTTCAACATGTCACCACTG 2220
Db      2161 CCAGACAAACCATGAGATTTTCCACGTTACCTATTTTCTTCAACATGTCACCACTG 2220
Qy      2221 CTCTTCTATAGAAAAGTATTTTTCATGAGTTAGGCAATTTTGAAGACAGTAGA 2280
Db      2221 CTCTTCTATAGAAAAGTATTTTTCATGAGTTAGGCAATTTTGAAGACAGTAGA 2280
Qy      2281 CAATTTGTCTATTTTCACTACTACACACTTACATTAACAGGTTGATTTTGAAGTTC 2340
Db      2281 CAATTTGTCTATTTTCACTACTACACACTTACATTAACAGGTTGATTTTGAAGTTC 2340
Qy      2341 ATCTGCTGCTAGATCCATGCTGATGATATGCTGACATACACCTACAGCAGATTC 2400
Db      2341 ATCTGCTGCTAGATCCATGCTGATGATATGCTGACATACACCTACAGCAGATTC 2400
Qy      2401 TCATCAAAAGTGAATGTGAGTGTGATGCTTCAGACTCTGATGATGATTTTGAAGTTC 2460
Db      2401 TCATCAAAAGTGAATGTGAGTGTGATGCTTCAGACTCTGATGATGATTTTGAAGTTC 2460
Qy      2461 TCAGTGTATATGTTGAGTAACTTTTGAACCCAGATTTGATCACTGAGGTTGAGC 2520
Db      2461 TCAGTGTATATGTTGAGTAACTTTTGAACCCAGATTTGATCACTGAGGTTGAGC 2520
Qy      2521 CATCTGTTTTCACGCTGCTGATACAGGATCACTTTCTTACCTTTTGAAGGTTTCA 2580
Db      2521 CATCTGTTTTCACGCTGCTGATACAGGATCACTTTCTTACCTTTTGAAGGTTTCA 2580
Qy      2581 GTGAATTTTCACTGTGTTCTGATGATGATGTTTGTTCATCTCTGTCACCTCAGCC 2640
Db      2581 GTGAATTTTCACTGTGTTCTGATGATGATGTTTGTTCATCTCTGTCACCTCAGCC 2640
Qy      2641 CCAGATTAAGGGGCTCATGCTGATGCTGCTTAATTTTCTGTAAGTTGGGGCTCC 2700
Db      2641 CCAGATTAAGGGGCTCATGCTGATGCTGCTTAATTTTCTGTAAGTTGGGGCTCC 2700
Qy      2701 TCCGCCCCCCTGTAAGAGTGTGTAAGCTTACAGGACACACTGATGATGCTG 2760
Db      2701 TCCGCCCCCCTGTAAGAGTGTGTAAGCTTACAGGACACACTGATGATGCTG 2760
Qy      2761 GGGGTGAAGTCAAGTGTGACAGCAGCTACAGCTGAGCTGAGCTTCAAGCAG 2820
Db      2761 GGGGTGAAGTCAAGTGTGACAGCAGCTACAGCTGAGCTGAGCTTCAAGCAG 2820
Qy      2821 GTTACTGAAGTCTGATCTTCTTGTCTTGTCTTGTGCTGCTG 2867
Db      2821 GTTACTGAAGTCTGATCTTCTTGTCTTGTCTTGTGCTGCTG 2867

```

RESULT 2

AA290578 standard; cDNA; 2610 BP.

AA290578;

19-JUN-2000 (first entry)

Human death inducer-obliiterator 1 (DIO-1) polypeptide encoding cDNA.

XX

Death inducer-obliiterator 1; DIO-1; cell death; cancer; tumour; human; autoimmune disease; cytostatic; immunosuppressive; antidiabetic; antirheumatic; antiinflammatory; antiproliferative; ss.

Homo sapiens.

Key	Location/Qualifiers
CDS	265..1953
FT	/tag= a
FT	/product= "DIO-1"

MO200015787-A1.

23-MAR-2000.

10-SEP-1999; 99WO-GB003019.

10-SEP-1998; 98SR-00003069.

17-SEP-1998; 98US-0100873P.

(CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.

(BANU/) BANNERMAN D G.

Alonso CM, Domingo DG, Grandien A, Leonardo B, Martinez P;

MPI; 2000-271426/23.

P-PSDB; AAY67579.

New DNA encoding human and murine death inducer-obliiterator 1 polypeptides, useful in the treatment of cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign tumors, malignant tumors and hyperproliferative skin disorders.

Claim 1; Fig 1A; 27pp; English.

The invention provides nucleic acids encoding the human and murine death inducer-obliiterator 1 (DIO-1) polypeptides. The polypeptides can be expressed by standard recombinant methodology. The DIO-1 polypeptides, agonists and antagonists are used as a medicament for treating diseases characterized by an alteration in cell death or by hyperproliferation, e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign tumors, malignant tumors or hyperproliferative skin disorders. They are also useful in the treatment of metabolic, proliferative or inflammatory conditions. The present sequence represents a cDNA encoding the human DIO-1 polypeptide

SQ Sequence 2610 BP; 644 A; 667 C; 734 G; 565 T; 0 U; 0 Other;

Query Match 35.7%; Score 1023.6; DB 3; Length 2610;

Best Local Similarity 71.9%; Pred. No. 3.3e-309; Matches 1494; Conservative 0; Mismatches 534; Indels 50; Gaps 10;

```

Qy      1 TCCGTGTAGCTCTGGAATGCTGCGGATCCCGCGCCGCGGAGCTTGTAAAG 60
Db      21 TCCGCGGCGCTTGCGGGAATGCTGCGGATCCCGGAGCTTGTAAAG 80
Qy      61 CAGTCCGACGTGCTTGTGTTGCAAGCTCCGGAATCTTCTATGATGAAGT 120
Db      81 GAAAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 140
Qy      121 ACTGACAAAGTGGGTGAG--GCTTGGCGCTGCTGAC-----TGGCC 166
Db      141 GGTGACAAAGTGTGTTGTGAGCTGCGTGTCTGTCGACCAAGGTTGCTG 200
Qy      167 CAGGTCTATTAATTTTATGAGATTTCCAGCCAAAGTTTCCAGTTTCAATTT 226
Db      201 AGGGTTTTTGTGTTATTTAGATTT--CAGGGAAGTGTCCAAAGTTTCAAGTT 259
Qy      227 CAGGTATGATGATTAAGGCACTGAGCAATGAGGACACCAAGGCTATCAAA 286
Db      260 CAGGTATGATGATTAAGGCACTGAGCAATGAGGACACCAAGGCTATCAAA 319
Qy      287 CCAGTAAAGATTTCAAGAAACCTGGGGTTTTTGAAGAAACCAAGATTTCCAA 346

```

```
Db 320 CCAGCAAGAGTTCAGAAAAATGGGGTTTTCAGAGACATATCCGCAAGCCAGAGG 379
Qy 347 GTGACAGAGACACGAGGCGGACCCAGTGAAGCAGCAAC-----ACAGCAGCATTAAC 400
Db 380 GCGCAGGGGAGAGCGGAGGCTGACCCATGAGGCGGCAACCCCAAGCAGCAGCTGGGCC 439
Qy 401 TCTCCCTGCGCGCAGTGAAGGCAACCAAACTGATGAGAGGTTAGAGATTCTTA 460
Db 440 TGTCCCTGCGCGCAGTGAAGGCAAGCCCAAGCAGCATGAGCGCTGAGGAGCTTCTGA 499
Qy 461 CCAAGGTTGCGCGCAGGAGAAAAAGATGCGCGGTGTCCTTGAAGATTCCAGTGAAC 520
Db 500 CCAATGCGCGCGCGCGCGCAGAGAGACATGCTGTCTCTCTGAGAGATTGTGTGAGAC 559
Qy 521 CCAATCTTCAACAGTCACTGATGTGAGACAGCTTCCGAGGGGAGCGTTGAAGACAGT 580
Db 560 CCAACTCTGCGCGCGCAGACCGCGAGACAGCTTCCGAGGGGAGCGCTGAGAAAGCGCTT 619
Qy 581 CTGAGATCAGAAAGTGGCCCTGTATCTGACTCTTA--GGGAAAGACATCTGCTCTT 637
Db 620 CTGAGACAGAAAGCGGCCCGCCAGTCTGCTTCCAGCTGTGAGAGAGACAGCAGCTCTT 679
Qy 638 CTGAAAAGGCAAAAGAGGTGAAGAGAAAGACACCTGACAGTGAAGTGAAGTGGCC 697
Db 680 CTGAAAAGGTGAAGAGAGAGAGTACACAGATGACCTCCGATGATGACAGCAGATGGCC 739
Qy 698 TTACGTTGAAGAACTTCAGAACCCCTTCGAGAAAGGAGAGAGCAAGCACTGTGAGA 757
Db 740 TGACCTTGAAGAAAGCTTCAGAAATGCGCTTCGAGAAAGGAGAGAGAGCCCACTGAGA 799
Qy 758 GGTCCCTGAGAGGCAAGTGAATGCGCTGAGAGAGAGAGTGCAGAGAGAAATTCGCCG 817
Db 800 GGGCCCTGAAAGGGATCCAGAGTCCCTGCGGAGAAAGCGCGAGAGGGTCTCCGCGCG 859
Qy 818 AAACGTGAGAGTCCAAATAGGCAAGTGCAGAGCAAGACCTCTGTATGACAGAGAC 877
Db 860 AGACTGTGGGCTCCAGAGGCAAGTGAACATGTTGAGAGGCGCTCCGCACTAGACAGAGAC 919
Qy 878 CTGAGGCTAGTCAAGGAGCAAGTGTCCAGTGAAGACAGATGACATAGAAAATCAAGTTGG 937
Db 920 CCGAAGACGATCAGGGGGGTGTGTCCAGGCGTGGAAAGATGACAGAGAGATTAAGTTGG 979
Qy 938 AAAGGAGGCGCATCTAGGGAATTAAGAGAAAAACCCAGGAGACGGGCAAAACCAAGC 997
Db 980 AGGGAAGGCGGCTCAGAGACATCAAAAGATGAGAGACCTTGAGACTTGGCCAGCGAAGC 1039
Qy 998 CTGAGTGAAGGTTAAGACCCCAATGCGCTGTACTGCAATCGCGGCGCACTCAACAAC 1057
Db 1040 CTGAATGAGAGGTTAAGACCCCAAGCGCTGTATTTGCGCGCAAGCTCAACMACA 1099
Qy 1058 ACAGGTTATGATCTGCTGTGATCCGATGAGAGTGGTTCATGAGTGAAGTGTGGTGA 1117
Db 1100 ACAGGTTATGATTTGCTGTGACCGCTGTGAAGAAATGGTTTCAATGCGATTTGTGGGA 1159
Qy 1118 TTTCTGAGGCGCGGAGCGGCTCTGTGAAGAGACGGGAGAGATCAATCTGCCCCAATT 1177
Db 1160 TTTCTGAGGCTCGAGGAGGCTTTTGTGAAGAAAGAGAGGAGAACTATATCTGCCCCA 1219
Qy 1178 GCACCATTTTGAAGTGAAGATGAGACAAACGATAGCGCCCAATGAGCAGAGCTCTG 1237
Db 1220 GCACCATTTCTGCAATGAGATGAGATGAGATCTATTCAGAAAAGGCAATCAGCAGAGAGCTA 1279
Qy 1238 GGTGCAATCTGTGGGTCTGATGAGCAGACAGCTGACAGATGAGGAGCAGTGAAGCAGA 1297
Db 1280 AATGAGACCTGAGAGATGCTGATGAGCAGATTTGTAAGATTAAGAACTAATAGAGCAGA 1339
Qy 1298 AGTCCGAGAAAGACCAAGGCGATTAAGGGTATGAGATGAGAGGCAACCAACCCAGCGCA 1357
Db 1340 AGTCTAGGAAAGACCAAGGAGTAAAGGGTATGAGAACTCAATTCCAATGTGGA 1399
Qy 1358 AGAAAAACTCAAGATATTTCCAGCGTGTGTAGAGGCTCTGTGATTAATGATTTG 1417
```

```
Db 1400 AGAAGAACTCAAGATCTTCCAGCCTGTGTATAGAGCGCGCTGTGCTCAAAATGATTTG 1459
Qy 1418 GCCCTGGGTGTTCAAGTATGACAGCCTGACTCTGTGTATTTGACGTATGACTGCAATTC 1477
Db 1460 GCCCGGGTGTCTACCTGTGCGGAGCGGACCTGCTGTATCTGCACTATATGACTGTATCC 1519
Qy 1478 TCAAAACAGCAGCAGCTACATGAGATTTCTAAGTTCAAGTAAAGAACAAAACAAAC 1537
Db 1520 TCAAAACAGCAGCAGCAGCAATGAGATTTCTAAGCTCAAGTAAAGAACAAAGCCAAAGC 1579
Qy 1538 CCAAGGAAAGGTCAAGACAGAGCCAGAAAAGTTCACTTTCAAAATGACAGTGTTCAG 1597
Db 1580 CTAAAGAAAGATGAAAGATGAAAGCCAGAGAGCCAGATCTTCCAAAATGCGGTGCTCAG 1639
Qy 1598 TGGGATTAATAATCTTTCTGTGACAAAGAGCTAGCCTCAGAGAAAGAGAAACCCAG 1657
Db 1640 CAGGTATTAATAATCTTTCTGTGACAAAGAGCAGCTCCAGAAAAGAAAGAGACACAG 1699
Qy 1658 TGAAGAA---AGTATGTGGCTTCCAGAGTGAAGATTTGAGAAAGAGAGAGCGCTGTG 1714
Db 1700 TGAAGAAAGCAGTGTGTCTCTGCGGAGTGAAGACTCGGAGAGAGAGAGAGCTTGTG 1759
Qy 1715 AGAGACGACACACATCTGAGGCAAGTGAACAACATAATGCTGTGAAGCCAGAGAGC 1774
Db 1760 AGAGACGACAGCCCTGTGTGTGGGAGAGATCAATTAACATGACAGTAAAGCCAGAAAAG 1819
Qy 1775 CAGAGAGCCCACTGCACTGTGCCACCCCTATTGATTAATGATATCACTCAACCCAAAG 1834
Db 1820 CTG-----CTGCTCCCTGCGCGCTGCACTGTGTATTAATGATATCACTCAAGG 1870
Qy 1835 CTGCTTCCAGGCGCCCTCCCATCATCT-----GGGTGGCTGCTGAGGCTGTCTAGAGC 1889
Db 1871 TTGGCTCTTGACACCTCTCCGTTCTTTCTGATATGACATCCCTGAGGCTGTCTCAGAGC 1930
Qy 1890 CAG-----AGTCTGGGTGTCTGTGTGCTGATAGTAGCCAGAGC---TCACTGCCAGC 1940
Db 1931 TGGAGTTGACAGCTTGTGTATTAAGCTGATCAAGACACCGGTGACCATCAGCGGGAAG 1990
Qy 1941 CAGAGCAGATATCAAGATGCTCTGTGAGACCCAGATGTTTCTGCTAGCCTGTGAGAGCT 2000
Db 1991 CAGAGCCATGTCCAGATATGCTCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2050
Qy 2001 CTCTGGGTGTTCTTAAGAGCTGTGTAGGCTCATGT 2038
Db 2051 TCTGTCACTGTTTCCAAAGCTGAACCTCATGCT 2088

RESULT 3
AAH15102
ID AAH15102 standard; cDNA, 2772 BP.
XX
AC AAH15102;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human CDNA sequence SEQ ID NO:13129.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000BP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
```

PA (HELI-) HELIX RES INST.

XX Oca T, Iacobi T, Nishikawa T, Hayaehi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.

DR Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.

PS Claim 8; SEQ ID NO 13129; 2537bp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention.

XX Sequence 2772 BP; 699 A; 691 C; 765 G; 617 T; 0 U; 0 Other;

XX Query Match 35.3%; Score 1013; DB 4; Length 2772;

XX Best Local Similarity 73.4%; Pred No. 7.4e-306;

XX Matches 1413; Conservative 0; Mismatches 475; Indels 36; Gaps 8;

DB 141 GCTTGCCGCTGCTTGAACCTGCGCCCAAGCTATATTATATGAGATTTCACGCCAA 200
QY
DB 211 GCTTGACCCCAAGGTTTGTCTGCGAGGTTTGTGTTATTTAGATTTCACGAGAA 269
QY
DB 201 AGCTTCCAGCTTTCAAGTTTGGACAGATATGATTAAGGACCTTGACATGA 260
QY
DB 270 AAGTGTCCAAAGCTTTCAAGTTTGGACAGATATGATTAAGGACCTTGACATGA 329
QY
DB 261 GGAAGCACCAGGAGGATCAAAACCCAGAGTAAAGAGTTCAAGAAACCTGGGATTTCG 320
QY
DB 330 GAGAGGACCTTAAGGACATCAAAACCCAGAGTAAAGAGTTCAAGAAACCTGGGATTTCG 389
QY
DB 331 AAGAACAGATATGCGCAAAAGGAGTCAAGAGACACAGAGGCGAGACCCAGTAGCA 380
QY
DB 390 AAGAACAGATATGCGCAAAAGGAGTCAAGAGACACAGAGGCGAGACCCAGTAGCA 449
QY
DB 381 GCAACCC-----ACAGCAGATTAACCTTCCCTGCGCCGCGAGTGAACCGCAACAAAG 434
QY
DB 450 GCCACCCCGACAGCAGCAGCTGGGCTGTGCTCCGCGCGCAGTGGAGGAGCGCCCAAGG 509
QY
DB 435 TACTGAGAGGAGTGAAGAGTTTCTTAACCAAGGTTGGCGCGCGAGGAGAAAGATGTGCC 494
QY
DB 510 CACTAGGCGCGTGAAGCAATCTGACCAATTGGCGCGCGCGCGAGGAGGAGCATGTGCC 569
QY
DB 495 GGTGTCCCTGAGAGATTCCAGTAGGCGCCACATCTTCCAGATCACTGATGTGAGACAG 554
QY
DB 570 TGTCTCCCTGAGAGATTCTGTGTGAGCCCAAGTCTGCGCCGCGACAGAGCGGAGACAG 629
QY
DB 555 TTCCGAGGAGAGCGTTGAAAGAGATTCTGAGATCAGAGTGGCTGTATCTGATCTCTT 614

DB 630 CTCGAGGCGACGCTGGAAGAGCGCTTCTGAGACCAAGAGCGGCGCCCAAGTCTGCTTCAC 689
QY
DB 615 A---GGGAAAGACATCTGCGCTCTTCTGAAAGGCAAGAGAGTGAAGAGAGAGA 671
QY
DB 690 AGCTGTGAAGAAAGCAGACCGCTCTTCTGAAAGGTAAGAGGAGGATGACCAAGATGA 749
QY
DB 672 CACCTCTGACATGACAGATGAGTGGCTTACGTTGAAGAACTTCAGAACCGCTCGGAG 731
QY
DB 750 CACCTCTGACATGACAGATGAGTGGCTTACGTTGAAGAACTTCAGAACCGCTCGGAG 809
QY
DB 732 AAGCGAGAGCAAGAACTGTGAGAGTCCCTGAGAGGAGTCAAGATGCGCTGAGAA 791
QY
DB 810 GAGCGGAGAACAGAGACCCCACTGAGAGGCGCTTGAAGAGATCCAGATGCGCTGAGAA 869
QY
DB 792 GAGCGGAGAGAGAGATTTCTGCGAAACTGAGAGTTCCAATAGCAGTGCAGCA 851
QY
DB 870 GAGCGCGGAGAGAGGATCCGCGCAGAGCTGTGGCTCCGAGCGCAGTACCTGTGGA 929
QY
DB 852 GAGCAGACCTCTCTTGAACAGAGGCTGAGAGCTTACGAGGACAGTGTCCAGTCA 911
QY
DB 930 GGGCGTCTGCGCCAGTAAAGAGAGGCGCCAGAAAGATCAGAGGAGTGTGTCCAGGCTGG 989
QY
DB 912 GACAGATGACATGAAATCAATTGGAAGGAGAGGCACTCAGGAAATACAGAGAA 971
QY
DB 990 GAAAGATGACAGAGAGTAAAGTTGAGAGGAAAGCGCGCTCAGACATCAAGATGAG 1049
QY
DB 972 CCGCAGAGAGCGGCGAAACCAAGCGCTGAGTGAAGTTTACGACCCCAATGCTCTGTA 1031
QY
DB 1050 GCGTGAAGACTTGGGCGCAGCCGAAAGCTGAATGAGAGGTTACGACCCCAAGCGCTGTA 1109
QY
DB 1032 CTGCACTGCGCGCAGCTTCAACAACAAGTTTATGATCTGTGAGATGAGTGTGAGCA 1091
QY
DB 1110 TTGCAATTTGCGCGACCTTCAACAACAAGTTTATGATGTTGTGACCGCTGTGAGA 1169
QY
DB 1092 GTGGTTCAATGAGATGATGTGGGTATTTCTGAGCGCGGAGGCGGCTCTGAAAGAA 1151
QY
DB 1170 ATGTTTCAATGAGATGATGTGGGTATTTCTGAGCGCTGAGGAGGCTTTTGAAGAA 1229
QY
DB 1152 CGGGAAGACATACATCTGCGCAATTTGACCATTTTTCAGAGTCAAGATGAGACAAACG 1211
QY
DB 1230 TGGGAAGACATATCTGCGCAATTTGACCATTTTTCAGAGTCAAGATGAGACATCTTC 1289
QY
DB 1212 TAGCGCCACATGATGACAGAGCTCTGGGTGACATCTGTGGGTGCTATATGACAGCTG 1271
QY
DB 1290 AGAAACGAGATCAGACAGAGAGCTAAATGAGACCTGAGATGCTGATGAGCACCGATTG 1349
QY
DB 1272 CACAGCATAGGAGCAGTGAAGCAGAGATCCGAGAAAGACAGAGGACATAAGGATAGAT 1331
QY
DB 1350 TACAGATATGAGAACATAGAGCAGAGATCTTACGAAAGACAGAGGATTAAGGATAGAT 1409
QY
DB 1332 TGAAGAGCAGCAAAACCCAGCGGCAAGAAACTCAAGATATTCAGCTGTCTGTAGA 1391
QY
DB 1410 TGAAGAAAGCTGGAATTCGAAGTGGCAAGAAACTCAAGATATTCAGCTGTGTAGA 1469
QY
DB 1392 GGTCTGCTGTCTCTTAATGATTTGGCTGTGGTGTTCAGATGATGACAGCTGTAGCTC 1451
QY
DB 1470 GGTCTGCTGTCTCTTAATGATTTGGCTGTGGTGTTCAGATGATGACAGCTGTAGCTC 1529
QY
DB 1452 TGTGTATTCAGATTAATGATGATGATTTCTCAACACGAGCAGTATACCTAGATTTCTAAG 1511
QY
DB 1530 GGTGTATTCAGATTAATGATGATGATTTCTCAACACGAGCAGTATACCTAGATTTCTAAG 1589
QY
DB 1512 TTCAAGTAAGAACCAAAACCAAAACCCAGAGAAAGTTCAGAGCAGAAAGTT 1571
QY
DB 1590 CTCAGGTAAGAACCAAAACCCAGAGAAAGTTCAGAGCAGAAAGTT 1649
QY
DB 1572 CAGTCTTCCAAATGATGATGATTTCTCAAGTGGAGATTAATTTCTTCTGTGACAAAGAGCT 1631
QY
DB 1650 CAGTCTTCCAAATGATGATGATTTCTCAAGTGGAGATTAATTTCTTCTGTGACAAAGAGCT 1709
QY
DB 1632 AGGCTCAGAGAAAGGAAACCCAGTGAAGAA---AGTATGCTGGCTTCCAGAGAGTGA 1688
QY
DB 1710 AGCTCAGAGAAAGGAAACCCAGTGAAGAAAGTGTGTGTCTCTGCGCGAGTGA 1769

QY 1689 GACTTCTGGAGAGGAGGAGCCTGTGAGAGCAGCAACCATTCTGGGAGAGTACCA 1748
 DB 1770 AGCACTGGGAGAGGAGGAGCCTGTGAGAGCAGCAACCATTCTGGGAGAGTACCA 1829
 QY 1749 CTACATCTGTGAGAGGAGGAGCCTGTGAGAGCAGCAACCATTCTGGGAGAGT 1808
 DB 1830 TTACATCTGTGAGAGGAGGAGCCTGTGAGAGCAGCAACCATTCTGGGAGAGT 1880
 QY 1809 GAGTAAATGTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1863
 DB 1881 GATTAATGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1940
 QY 1864 GGTGCTCTGTGGGCTGTGATGAGCAG-----AGTCTGGGTTCTGTGATGATG 1918
 DB 1941 AGCATTCCCTGGGCTGTGATGAGCAGCTGGAGTTGATGATGATGATGATGAT 2000
 QY 1919 TAGCAGAGC-----TCACTGCAAGCAGAGATGATGATGATGATGATGATGAT 1974
 DB 2001 ACACCGGCTGACCATGAGCGGAGAGAGCCATGTCAGAGATGCTCTGCTGCT 2060
 QY 1975 GTGTTCTGTGAGCCTGTGATGAGCAGCTGTGATGATGATGATGATGATGATGAT 2034
 DB 2061 GTGTCATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2120
 QY 2035 ATGT 2038
 DB 2121 TGGT 2124
 RESULT 4
 ABV28029 ID ABV28029 standard; cDNA; 2085 BP.
 AC ABV28029;
 XX 16-SEP-2002 (first entry)
 DT Human prostate expression marker cDNA 28020.
 DE Human prostate expression marker cDNA 28020.
 XX Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX Homo sapiens.
 OS WO200160860-A2.
 PN 23-AUG-2001.
 DB 20-FEB-2001; 2001MO-US005171.
 PF 17-FEB-2000; 2000US-0183319P.
 PR 16-MAR-2000; 2000US-0189862P.
 PR 25-MAY-2000; 2000US-0207454P.
 PR 09-JUN-2000; 2000US-0211314P.
 PR 18-JUL-2000; 2000US-0219007P.
 PR 13-DEC-2000; 2000US-0255281P.
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX Schlegel R, Endege WO, Monahan JE;
 PI WPI; 2001-662795/76.
 DR Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX Claim 1; Page 5788; 11750p; English.
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether

CC a patient is afflicted with prostate cancer; (b) monitoring the
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
 CC
 SQ Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;
 Query Match 33.1%; Score 949.8; DB 5; Length 2085;
 Best Local Similarity 73.5%; Pred. No. 46-286;
 Matches 1884; Conservative 0; Mismatches 462; Indels 37; Gaps 12;
 1 TCCGTGTAGTCTGTGAAATGCTGCGGATCCCGCGCGCGGAGCTTGTAAAGAG 60
 DB 35 TCCGCGGGTTCGGGAAATGCTGCGAGACCTTAAGGCTGCGGAGCTTACTCAGG 94
 QY 61 CAGTCCCACTGCTCTTGTGTTTCAAGCTCCGAACTTCTCATTTAGTACTG 120
 DB 95 GAAAGAGCTTGTATATCTGATGATGATGATGATGATGATGATGATGATGAT 153
 QY 121 ACTGACACAGTGGGT--GAGGCTTGGCTCTGCTT-----ACCTGCCCC 167
 DB 154 GCTGACACAGT 213
 QY 168 AGGTCTATATTTTATGATGATTTTCAAGGATTTTCAAGCTTTCAGTTTGGAG 227
 DB 214 AGGTTTGT 271
 QY 228 AGGTATGATGATTAAGGACCTGAGATGATGATGATGATGATGATGATGATGAT 287
 DB 272 AGGTATGATGATTAAGGACCTGAGATGATGATGATGATGATGATGATGATGAT 331
 QY 288 CAGTAAAGATGATTAAGGACCTGAGATGATGATGATGATGATGATGATGATGAT 347
 DB 332 CAGCAAGATGATTAAGGACCTGAGATGATGATGATGATGATGATGATGATGAT 391
 QY 348 TGCAGAGACACGAGAGGAGACCCAGTGTAGAGCA-----ACCAAGAGATTA 402
 DB 392 CGCAGGAGACGAGAGGAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 451
 QY 403 TCCCTGCGCGGAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 462
 DB 452 TCCCTGCGCGGAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 511
 QY 463 AGGTTGCGCGGAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 522
 DB 512 ATTGCGCGGAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 571
 QY 523 ACATCTTCCACAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 582
 DB 572 AGTCTCTGCGCG--CAGCAGAGCGGAGCAGC--TCCGAGGCGAGCGTGTAAAG 629
 QY 583 GAGATCAGAGT 642
 DB 630 GAGACAGAGAGGCGCGCGCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 689
 QY 643 AAGGCAAAAG 702
 DB 690 AAGGTGAAG 748
 QY 703 TTGAAGAACTTCAAGACCGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 762
 DB 749 TTGAAGAGCTTCAAGATCG--CTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 807
 QY 763 CTGAG 822
 DB 808 CTGAAG 867
 QY 823 GGGAGTGTCAATATAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 882

Db	868	GTGGGCTCCGAGGCCAGTGAACA	CTGTGAAGGCGCTTCG	CCAGTAAAGCGAGGCCGAG	927			
Qy	883	GCTAGTCAGGGAC	CAGTGTCCCA	GTCAAGACAGATGA	CTATGA	AAATCATGTTGGAAAGG	942	
Db	928	AAACATCAAGGGGGTGTGTCTCC	AGGCTGGGAA	AGATGA	CAAGAGAGATGA	AGTTATGGACGGA	987	
Qy	943	AAAGGCACTCAGGGA	ATACAGAGGAA	CCCAAGGAA	AGCGGCA	AAACCAAGCTGAG	1007	
Db	988	AAAGGCGCTCAGSAC	ATCA	AAATATAGGAG	CGCTGAG	AGACTTGTGGCGAC	ACGAAAGCTGAA	1047
Qy	1003	TGTAGAGTTTACGAC	CCCAATGCGTGTAT	CTGCATCTGCGC	AGCTCA	CAACACAGS	1067	
Db	1048	TGTAGGGGTAAAGAC	CCCAACCGCTGTAT	GTGCA	TTTGCGC	AGCTTACACACAGS	1107	
Qy	1063	TTTATGATCTGCTGTGAT	CGGTGTAGAGAGTGGTT	CCATGTGATCTGTGTGGTAT	TTTTCT		1122	
Db	1108	TTTATGATTTTGTGTGTGA	CCGCTGTAAAGAT	TGGTTTCA	ATGGGATGTGTGTGGC	CA	TTTCT	1167
Qy	1123	GAGGCCGAGGGCGGCT	CTGTGAAAGAA	CGGGAAAG	ACTACATCTGCC	CAATTCAC	1187	
Db	1168	GAGGCTCAGAGGAGGCT	TTTGTGAAAGAA	TGGGAAAG	ACTATATCTGCC	CAATTCAC	1227	
Qy	1183	ATTTTGCAGGTGACAGAT	AGAGCA	CAACGGTACGGC	CA	CAATAGAGAGGACT	CTGGGTGC	1247
Db	1228	ATTCTGCAAGTGCAGAGAT	AGAGCTCTTC	TCAGAA	CGCGAGATCA	AGAGAGTAA	TG	1287
Qy	1243	AGATCTGTGGGTCTGAT	GGCA	CAGACTGCA	CAAGCAT	TAGGACAGT	AGAGAGTCC	1307
Db	1288	AGACTGTGAGATGCTGAT	GTGAC	CCGATTTGTA	CAAGTATAGAA	CAATAGAGCA	AGAGTCT	1347
Qy	1303	GGAGAGAC	CAGGCGATTA	AGGTTGAT	TGAGAGG	CGAGCA	ACCCAGCGCGAGAA	1367
Db	1348	AGCGAAGAC	CAAGGGATAA	AGGTGAGAT	TGAGAA	AGCTGCA	AAATCCAGTGGCAAGAG	1407
Qy	1363	AAACTCAGAGAT	ATTTCCAGCTGT	CGTAAGAGGCT	CCGTGGTCT	CCTTAAT	TGATTTGGCCCT	1427
Db	1408	AAACTCAAGAT	TTTCCAGCTGT	GTATAGAGG	GGCTGGTCT	CAAAATG	ATTTGGCCCC	1467
Qy	1423	GGGTGTCCAGTGTAG	CA	CAGGCTGAT	CTGTGTAT	ATTCAGT	TAATGACTGCA	1487
Db	1468	GGGTGTCTGTCA	GTGGCGGAGCCG	CACTCGGTAT	CTGCA	GTATATG	ATGATCTCAAA	1527
Qy	1483	CACGCAAGCTAC	CA	TGAGATTTCT	TAAAGTTCA	AGGTAAAGAA	CAAAATCCAA	1547
Db	1528	CACGCGAGGCA	CAATGA	AGTTCTTA	AGCTCAGGTAA	AGAA	CAGAGCCAAAGCTTAA	1587
Qy	1543	GAAAGGCTCAGAG	CGAAGCGA	AAAGTTCA	GTCTTCCAA	ATATG	CAAGTTCAGGTGGG	1607
Db	1588	GAAAGATGA	AGATATAG	CGAAGAG	AGCCACTTTC	GAAATG	CGGTCTCAGGCAAGT	1647
Qy	1603	ATTAAATCTCTT	CTGTGCA	CA	MAGACTAG	CGTCA	AGAGAAAGGAA	1667
Db	1648	ATTAAATCTCTT	CTGTGCA	CA	MAGAGCA	CAAGCTCA	AGAGAA	1707
Qy	1663	AA---AGTGA	TGCTGCTTCC	AGAGTGA	CACTTCTGG	AGAGAGG	CGCTGTGAGGC	1717
Db	1708	AAAGCAGTGTGTGTCT	CTGTGCGGGAGT	TAAGCA	CTCGGAA	AGAGAG	CAAGCTGTGAGAC	1767
Qy	1720	AGCA	CACATCTGTGGCA	AGTGA	CCACA	CTCAT	TGCTGTAGAC	1777
Db	1768	AGCA	AGCGGTGTGTGGCA	AGGATCA	ATTA	CAATG	CAATTAAC	1827
Qy	1780	AAAGCCAT	GTGCACTTGCCCA	CCCTAT	TGAGTAA	ATGAC	GTATCA	1837
Db	1826	-----CTG	CTCCCTGCGGCA	CTGTGTGTAT	ATATG	ATATCA	CTTAC	1877
Qy	1840	TTCCAGG	CCCTCCCA	TCATCT	1862			
Db	1879	CTCCAGAC	CCCTCCGTTCTT	1901				

ABV27935
ID ABV27935 standard; cDNA; 2085 BP.
AC
XX ABV27935;
XX
DT 16-SBP-2002 (first entry)
XX
DB Human prostate expression marker cDNA 27926.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PM W0200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PA Schlegel R, Endege WO, Monahan JB;
XX
PI WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.

Claim 1; Page 5757-5758; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising
a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
specification or its complement. (I) is useful for: (a) assessing whether
a patient is afflicted with prostate cancer; (b) monitoring the efficacy
of progression of prostate cancer in a patient; (c) assessing the efficacy
of a test compound to inhibit prostate cancer in a patient; (d) assessing
the efficacy of a therapy for inhibiting prostate cancer in a patient;
(e) selecting a composition for inhibiting prostate cancer in a patient;
(f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;

Query Match 33.1%; Score 949.8; DB 5; Length 2085;
Best Local Similarity 73.5%; Pred. No. 4e-286;
Matches 1384; Conservative 0; Mismatches 462; Indels 37; Gaps 12;

1 TCCTGTGAGTCTGGAAATGCTGCCGAATCCCGCGCGCGGAGCTGTTAAGAG 60
Db 35 TCCGCGGCGTTTCGGGAATATGCTGCAGACCCCTTAGAGGCTTGCGAGACTTACACGG 94
Qy 61 CAGTCCCACACTGTCCCTTGTGTTTTCGAGAAGCTCCGNAATCTTCATTGATGAATCTG 120
Db 95 GAACAGCTCTAGTATCTAAGTCTTGAATAATC-GAAGCTGTACTCTGTAACAATG 153
Qy 121 ACTGACAACAGTGGGGT-GAGGCTTGCGCTCTGCTTG-----ACCTGGCCCC 167
Db 154 GCTCAACAACAGTGTGTGTGATGAGCTGGCTGTCTGCTTGGAACCAAGAGTTCGCTGCC 213
Qy 168 AGGCTCATTAATTTATATGAGATTTCCAGCCAAGTTTTCCAACTTCAATTTTGGAGC 227
Db 214 AGGGTATTGTGTGATTTAGGATTT-CAGGGAAAAGTGTCCAACTTTCAG-TGGGAGC 271


```

Qy 228 AGGTATGATGATTAAGGCACTGAGCAATGAGGAAGCAACCAAGCTATCAAAACCAC 287
Db 272 AGGTATGAGAGCAAAAGCCGACCCGAGCAATGAGAGGCACTTAAGGCAATCAAAACCAC 331
Qy 288 CAGTAAGAGATTCAGAAAACTGGGGTTTTCGAAGAACCAAGATTGCAAACTGAGGG 347
Db 332 CAGCAAGAGATTCAGAAAACTGGGGTTTTCGAAGAACCACTATTCGCAAGCCAGAGGG 391
Qy 348 TGCAGAGACGAGAGGCGAGCCCACTGAGCAGCA-----ACCAAGACAGCATTAACCTC 402
Db 392 CGCAGGGAGCGCGAGGCTGACCCCACTGAGCCGACCCCAAGCAGCAGCACTGAGGCTG 451
Qy 403 TCCCTGCGCCGCAATGAGACGCAACCAAAAGTACTGAGAGGATTCAGTGAAGCTTAC 462
Db 452 TCCCTGCGCGCAATGAGAGGCGAGCCCAAGCGCACTGAGAGGCGAGCAATTCCTGAC 511
Qy 463 ACAGTTGCGCGCCGAGGGAAAAAGAAATGTCGCGGTGTCCTGAGAGATTCCAGTGAAGCC 522
Db 512 ATTGCGGCGCGCCGCGAGAGAGAGCATGCTGTCTCTCGAGGATTCGTGAGGCC 571
Qy 523 ACATCTTCACAGTCACTGATGTGAGACAGCTTCCAGGGAGGCTTGAAGCATTTCT 582
Db 572 ACCTCTGCCCCG--CAGCAGAGCCGAGACAGC--TCCGAGGCGAGCGTGAAGCGCTTCT 629
Qy 583 GAGATCAAGATGCGCCCTGTATCTGACTCCCTTGAAGGAAGAACTCCGCGCTTCTGAA 642
Db 630 GAGATCAAGAGCGGCGCCGAGTCTCTTCCAGATGTGAAGAGCAAGCATCTTCTGA 689
Qy 643 AAGCAAAAGAGGTGAAGAGAAAGAACACTCTGACAGTGAAGTATGAGCTTACG 702
Db 690 AAGGTGAAGAGAGGGATGACCAAGATGACCTCCGATGATGACAGGATGAG--CTGACC 748
Qy 703 TTGAAGAACTTCAGAACCGCTTTCGAGAAAGCGAGCAAGAACTGTGAGAGGTCC 762
Db 749 TTGAAGAGCTTCAGAAATCG--CTTCGACAGAAAGCGGAAACAGAGCCCACTGAGAGGCC 807
Qy 763 CTGAGAGGCACTCAAGATTCGCTGAGAGAAAGGCGAGAGGAAGATTTGCCGAAC 822
Db 808 CTGAGAGGATTCAGAGATTCGCTGAGAGAAAGCGCCGAGAGAGGTCCGCGCAACT 867
Qy 823 GGGAGTGTCCAAATAGGCAGTGCCGAGAGGACAGACCTCTCTGTAGCAGAGGCTGAG 882
Db 868 GTGGGCTCCGAGGCGCAAGTGAACATGTGAGAGGCGCTCTGCCAGTAAACAAGAGCCGAG 927
Qy 883 CCTAGTCAAGGACCAAGTGTCCCACTCAGAGACAGTGAACATTAAGAAATCAAGTTGAGAGG 942
Db 928 AACCATCAGGGGGGTGTGTCCAGGCTGGAAGATGACAGAGAGTAAATTGAGAGGGA 987
Qy 943 AAGCGCACTCAGGGAAATTAAGAGGAAAAACCCGAGGAAAGGCGCAACCAAGGCTGAG 1002
Db 988 AAGGCGGCTCAGAGACATCAAGATTAAGAGGCTGAGAACTTGGGCGCAAGCCGAGCTGAA 1047
Qy 1003 TGTGAGGTTTACGACCCCAATGCTCTGTACTGATCTGCGCCGACAGCTCACAACCAAG 1062
Db 1048 TGTGAGGTTTACGACCCCAAGGCTCTGTACTGATTTGCGCCGACAGCTCACAACCAAG 1107
Qy 1063 TTTATGATCTGCTGTGATCGGTGAGAGAGTGTTCATGTGACTGTGTGGGTATTTCT 1122
Db 1108 TTTATGATTTCTGTGACCGCTGTGAAAGATGTGTTCAATGCGATGTGTGGGCAATTTCT 1167
Qy 1123 GAGGCGGAGGGCGGCTCTGGAAGAAAGAACGGGAAAGACTACATCTGCGCAAAATTCGACC 1182
Db 1168 GAGGCTCAGAGGAGGCTTTTGAAGAAAGAAATGGGAAAGACTAATCTGCGCAAACTGACAC 1227
Qy 1183 ATTTTGAAGTGCAGAGATGAGCAAAAGGTAGCGCCACCAATGAGCAGAGACTTGGGTGC 1242
Db 1228 ATTCTGCAAGTGCAGAGATGAGCTCATTCAGAAACGGGAGATCAGCAGAGAACTAAATGG 1287
Qy 1243 AGATCTGTGGGTGCTGATGAGCAGACTGCAAGCATTAAGGACAGTGAAGCAAGATTC 1302
Db 1288 AGACTGTGAGATGCTGATGAGCAGACTTGTACAAAGTATTAAGAAACAATGAGCAGAGTCT 1347

```

```

Qy 1303 GAGAGAGCAGGCGCATTAAGGGTAGATTGAGAGGCAAGCAAAACCCAGCGGCAAGAA 1362
Db 1348 AGCGAGACCAAGGATTAAGGGTAGATTGAGAGGCAAGCAAAATCCAAAGTGGCAAGAG 1407
Qy 1363 AAACCTCAAGATTAATCCAGCTGTCTGTAGAGGCTCTGGTCTCTTAATGATTTGGCCCT 1422
Db 1408 AAACCTCAAGATCTTCCAGCTGTGTATGAGGCGCTGTGTCTCAAAATGATTTGGCCCT 1467
Qy 1423 GGGTGTCCAGTGAAGCAAGCCGACCTGCTGTGTATTTAGTACAGTAAATGACTGATTTCTCAA 1482
Db 1468 GGGTGTGTCACTGTGGCGAGCCGACCTGGGTGATGTGCACTTAATGACTGATTTCTCAA 1527
Qy 1483 CAGCAGCAGCTACCATAGATTCTTAAGTTCAAGTAAAGAACAAACCAAAACCAAG 1542
Db 1528 CAGCCGCGAGGCAATGAAGTTTCTAAGTCAAGTAAAGAACAAAGCCAAAGCCTAA 1587
Qy 1543 GAAAGGTCAAGACAGAAAGCCAGAAAAATTCAGTCTTCCAAATGCAATGCTTCAAGTGGGG 1602
Db 1588 GAAAGATGAAGATGAAGCCAGAGAGCCAGTCTTCCAAATGCGGTGCTCAGGCAAGT 1647
Qy 1603 ATTAATATCTCTCTGTGACAGAGACTAGCGTCAAGAGAAAGGAGAAACCCAGTGAAG 1662
Db 1648 ATTAATATCTCTCTGTGACAGAGAGACTAGCGTCAAGAGAAAGGAGCAAGTGTGAGAG 1707
Qy 1663 AA--AGTATGCTGCTTCCAGAGATGAGACTTCTGGAAGAGAGGCAAGCTGTGAGAGC 1719
Db 1708 AAGCAGTGTGTGTGCTTCCGCGAGGAGGACATCGGAGAGAGAGCAGTGTGTGAGAGC 1767
Qy 1720 AGCAGCAGCTCTGGGCAAGTGAACCACTAATGCTGTGAAGCCAGAGAGCCAGAG 1779
Db 1768 AGCAGCGCGTGTGGGCGAGGATCAAAATTAATGATGAGTAAAGCCAGAAAGACTG-- 1825
Qy 1780 AAGCCAGCTGCACTCTGCGCCCAACCTATTTAGTAAATGATGATCAACCAAGGCTGAC 1839
Db 1826 -----CTGCTCTCTGCGCGTCACTGTGTAAATGATGATCACTGAGGGTGGC 1878
Qy 1840 TTCCAGGCGCCCTCCCATCATCT 1862
Db 1879 CTCCTGAGACCCCTCCCGTTCTT 1901

RESULT 6
ABV22095
ID ABV22095 standard; cDNA; 2085 BP.
XX
AC ABV22095;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 22086.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189863P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX

```


QY 1780 AAGCCACTGACACTCTGCGCCACCCCTATTGAGTAATGTAAGTATCACCCAAAGGCTGGC 1839
DB 1826 -----CTGCTCCTCGCCCTGACTGTGTAATAATGATATATCACTAGGGGTTGGC 1878
QY 1840 TTTCCAGGCCCCCTCCATCATCT 1862
DB 1879 CTCCTGACCCCTCCGCTTCTTT 1901

RESULT 7
ABV22192
ID ABV22192 standard; cDNA, 2085 BP.
AC ABV22192;
XX
XX 13-SEP-2002 (first entry)
DT
DE Human prostate expression marker cDNA 22183.
XX
XX Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200160860-A2.
PN
XX
XX 23-AUG-2001.
PD
XX
XX 20-FEB-2001; 2001MO-US005171.
PF
XX
XX 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0239007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA
PI Schlegel R, Endege WO, Monahan JB;
XX
XX WPI; 2001-662795/76.
DR
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX
XX Claim 1; Page 3816; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
CC
XX
XX Sequence 2085 BP; 520 A; 526 C; 656 G; 378 T; 0 U; 5 Other;

Query Match 33.1%; Score 948.2; DB 5; Length 2085;
Best Local Similarity 73.4%; Pred. No. 1.3e-285;
Matches 1383; Conservative 0; Mismatches 463; Indels 37; Gaps 12;

QY 1 TCCGTGAGTCTTGAATAATGCTGCGGATCCGCGCGCGGAGCTTGTTTAAGAGG 60
DB 35 TCCGCGGCTTCGCGGAAATGCTGCGAGACCTTAGAGGCGCTGCGAGCTTACTCCAGG 94

QY 61 CAGTCCCACTGTCCTTGTGTTGTCGAAGCTCCGAAATCTTCTCATGTATGAATCTG 120
DB 95 GAACAGCCTTATGATATCTGATGTTGAAATATC-GAAGCTGTATCTCGTGAACAGT 153
QY 121 ACTGACAAACAGTGGGT--GAGCTTGGCCCTCTGCTTG-----ACCTGCCCC 167
DB 154 GCTGACAAACAGTGTGTGTGTGAGCTCGGCTGTCTCTGACCCAGAGGTTCTGTGCTC 213
QY 168 AGGTCTAATATTTATGTATGATATTTCCAGCCCAAGGTTTCCAGGTTTCAGTTTGGAC 227
DB 214 AAGGTTTGTGTGTATTTATGATTTT-CAGGAAAGTGTCCAGCTTTAG-TGTGAGC 271
QY 228 AGGTATGATGATTAAGGGAACCTGAGCAATGAGGAACCAAGGCTATCAAAACCCAC 287
DB 272 AGGTATGACACAAAGGCGAACCGAAGCAATGAGAGGCACTTAAGCATGAACCCAC 331
QY 288 CAGTAAGAGTTTCAGGAAACCTGCGGTTTTCGAAGAACCAAGATTGCCAAACGTGAGG 347
DB 332 CAGCAAAAGATTCAAGAAACATGCGGTTTTCGAAGACCACTATCGCCAAACGAGAGG 391
QY 348 TGCAGGAACACAGAGGCGAACCCCAAGTGAAGCA-----ACCAAGAGCATTAACCTC 402
DB 392 CGCAGGAGACCGGAGGCTGACCACTGAGCCGACCCCAAGAGCAGCAGCTTGAGGCTTG 451
QY 403 TCCCTGCGCGGAGTGAACGCAACCAAAACGTACTGAGAGGATGAAGATTCTTACC 462
DB 452 TCCCTGCGCGGAGTGAACGCAACCAAAACGTACTGAGAGGATGAAGATTCTTACC 511
QY 463 ACGGTTTCGCGCGGAGGAAAAAGATGTGCGGTTCTCTGAGAGATTCAAGTGAAGCC 522
DB 512 ATTGCGCGCGCGGAGGAGAGAGATGCTGCTCTCTGAGAGATTCTGAGGCCC 571
QY 523 ACATTTTCAAGTACATGATGTGAGACAGCTTCCGAGGAGGAGGTTGAAGCAAGTTCT 582
DB 572 ACGTCTGCCCCG-CACAGAGCCGCAAGCAAG-TCCGAGGAGGCTGGAAGAGCCCTTCT 629
QY 583 GAGATCAGAAAGTGGCCCTGTATCTGACTCTTGAAGAAAGAAATCTGCTCTTCTGAA 642
DB 630 GAGACCAAGAGCGGCTCCAGTCTGCTTCCACAGTGTGAAGAGACAGACAGCTCTTCTGA 689
QY 643 AAGCAAAAGAGGAGTGAAGAGAGAGAGAACCTCTGACAGTGAAGTATGAGCTTACG 702
DB 690 AAGTGAAGAGAGGAGGAGTGAACCAAGTACACCTCCGATGATGACAGCATAGG-CTGACC 748
QY 703 TTGAAGGAATCTTCAAGACCGCTTCTGAGAAAGGCAAGAGCAAGAACTCTGAGAGTCTC 762
DB 749 TTGAAGAGCTTCAAGAAATCG-CTTCCAGAGAAAGCGGAAACAGAGACCCCACTGAAGGCCC 807
QY 763 CTGAGAGGCAATCAAGATCGCTGAGAGAAAGGCGCAGAGAGAGAAATTTCTGCCGAACT 822
DB 808 CTGAAGGGAATCCAGAGTGGCTTCCGAGAGAGAGCCCGGAGAGAGGTTCCGCCAGACT 867
QY 823 GGGAGTGTCCAAATAGGAGATGCCAGAGAGACAGACCTTCTGTGAAGAGAGCTGAG 882
DB 868 GTGGGCTCGAGGCGCAAGTGAACCTGTGAGGCGGCTGCTGCCAGTGAAGAGAGCCGAG 927
QY 883 GCTAGTCAAGGACCAAGTGTCCAGTCAAGACAGATGACATGTGAAGAAATCAATTTGAAGG 942
DB 928 AACGATCAAGGGGTTGTGTCCAGGCTGGAAGATGACAGAGAGATGAAGTTGAAGGGA 987
QY 943 AAGGCACTCAAGGAAATACAGAGGAAACCCAGAGAAAGGAGGCAACCAAGCTGAG 1002
DB 988 AAGGCGCTCAGAGACATCAAGATGAGAGCTTGAAGACTTGGCGCAGACGAAAGCTGAA 1047
QY 1003 TGTGAGTTTACGACCCCAATGCTGTATGATCATCTGCGGAGCTTCAACAACAGG 1062
DB 1048 TGTGAGGTTTACGACCCCAAGGCTGTATGATGATTTGCGCGCAGACTTCAACAACAGG 1107
QY 1063 TTTATGATCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1122
DB 1108 TTTATGATTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1167
QY 1123 GAGGCGCAGAGGCGGCTCTCTGAAAGAAAGAGAGAGAGATCAATCTGCTCAATTTGACC 1182


```

QY 833 AAATAGGAGTGGCGAGGAGGACAGACCTCTGTGTAGCAGAGGCTGAGCTAGTCAG 892
DB 484 AGCCGAGTGCACCTGTGAGAGGCGCTCTGCTCCAGTACAGAGAGCCGAGAACATCAG 543
QY 893 GACCACTGTCCAGTCAAGACAGATGACATAGAAATCACTGGAGAGGAGGACTC 952
DB 544 GCGTTGTGTCCAGGCTGGAAAGATGACAGAGAGATGATGGAGGAAAGGCGCTC 603
QY 953 AGGAAATATCAGAGGAAACCCGAGGAGGCGGAAACCAAGCTGATGTGAGTT 1012
DB 604 AGGACATCAAGATGAGAGGCTTGAGACTTGGGCGAGCCGAGGCTTAATGTAGGGTT 663
QY 1013 AGGACCCCAATGCTGTATGACATCTGCGCGAGCTCAGACAAAGGTTTATGATCT 1072
DB 664 AGGACCCCAAGGCTGTATGATTTGCTCCGAGCTCAGACAAAGGTTTATGATTT 723
QY 1073 GCTGTGATCGGTGTGAGAGAGTGTTCATGATGATGATGATGATGATGATGATGAT 1132
DB 724 GCTGTGATCGGTGTGAGAGAGTGTTCATGATGATGATGATGATGATGATGATGATGAT 783
QY 1133 GCGGCTCTCTGAAAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1192
DB 784 GAGGCTTTTGGAAAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 843
QY 1193 TGAGAGATGAGACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1252
DB 844 TGAGAGATGAGACATCTATTCAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 903
QY 1253 GTGCTGATGAGCAGACATGACAGACATAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1312
DB 904 ATGCTGATGAGCAGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 963
QY 1313 AGGCAATTAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1372
DB 964 AAGGATTAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1023
QY 1373 TATTCAGGCT----- 1383
DB 1024 TCTTCAGGCTGTGTGCTGCGGCTGTGCTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1083
QY 1384 -----G 1384
DB 1084 AATATGCTGTGTGATGAGAGCATCTGCTTACACTTGTGACATGATGATGATGATGAT 1143
QY 1385 TCGTATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1444
DB 1144 TGATATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1203
QY 1445 CTGACTGTGTGTATGCTGATATGATGATGATGATGATGATGATGATGATGATGATGAT 1504
DB 1204 CCGACTGTGTGTATGCTGATATGATGATGATGATGATGATGATGATGATGATGATGAT 1263
QY 1505 TTCTTAAGTTCAAGTAAAGAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 1564
DB 1264 TTCTTAAGTTCAAGTAAAGAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 1323
QY 1565 AAAAGTTCAAGTTCTTCAAAATGAGTGTTCAGTGTGGGATTAATTAATCTTCTGTGACA 1624
DB 1324 AGAAGCCAGTCTTCCGAAATGCGGTGCTCAGGCGAGTATTAATTAATCTTCTGTGACA 1383
QY 1625 AGAAGTTCAAGTCTCAGAAAGGAGAAACCAAGTGAATAA---AGTATGCTGCTTCCA 1681
DB 1384 AGAAGTTCAAGTCTCAGAAAGGAGAAACCAAGTGAATAAAGGAGTGTGCTTGTGCGC 1443
QY 1682 GAGTGTGAGATTTCTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1741
DB 1444 GAGTGTGAGACACTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1503
QY 1742 ACACCACTTACATGCTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1801
DB 1504 ATCACTTACATGCTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1554
QY 1802 CCTATTGAGTAATGATGATGAT 1823

```

```

DB 1555 CACTGTGTATTAATTCACGAA 1576

RESULT 9
AAK52368
ID AAK52368 standard; cDNA; 7838 BP.
XX
AC AAK52368;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 913.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN MO200157190-R2.
XX
PD 09-AUG-2001.
XX
PE 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663551.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HVS8-) HXSEQ INC.
XX
PI Tang YT, Liu C, Dermanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
XX
PT P-PSDB; AAM79235.
XX
PS Nucleic acids encoding polypeptides with cytokine-like activities, useful
in diagnosis and gene therapy.
XX
PS Claim 1, Page 3007-3013; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 7838 BP; 1672 A; 2123 C; 2118 G; 1725 T; 0 U; 0 Other;
XX
Query Match 25.1%; Score 719.2; DB 4; Length 7838;
Best Local Similarity 70.7%; Pred. No. 2,1e-213;
Matches 1076; Conservative 0; Mismatches 323; Indels 123; Gaps 4;
QY 416 GTGAGCGCAACCAAGTACTGAGAGGATGATGATGATGATGATGATGATGATGATGAT 475
DB 599 GTGAGAGGCAACCAAGCAGCTGAGCGGCTGAGCAGTCTCTGACATGTCGCGCGCGCC 658

```

QY 476 GAGGAGAGAGAGATGTCGCTGTCCTGAGAGATTCAGTGAAGCCACATCTTCCACAG 535
 DB 659 GCGGAGAGAGAGAGATGTCGCTGTCCTGAGAGATTCAGTGAAGCCACATCTTCCACAG 718
 QY 536 TCACAGATGAGAGAGATTCGAGAGAGAGAGATTCAGTGAAGCCACATCTTCCACAG 595
 DB 719 CCACAGAGAGAGAGATTCGAGAGAGAGAGATTCAGTGAAGCCACATCTTCCACAG 778
 QY 596 GCCCTGATGATCTCTTA---GGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGAG 652
 DB 779 GCCCTGATGATCTCTTA---GGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAG 838
 QY 653 GAGTGAAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 712
 DB 839 GAGGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 898
 QY 713 TTCAAGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 772
 DB 899 TTCAAGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 958
 QY 773 GTCAAGATGATCTCTTA---GGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAG 832
 DB 959 TCAGAGATGATCTCTTA---GGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAG 1018
 QY 833 AAATAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 892
 DB 1019 AGGCGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1078
 QY 893 GACCAAGATGATCTCTTA---GGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAG 952
 DB 1079 GAGTGAAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1138
 QY 953 AGGAGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1012
 DB 1139 AGGAGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1198
 QY 1013 AGGAGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1072
 DB 1199 AGGAGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1258
 QY 1073 GCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1132
 DB 1259 GCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1318
 QY 1133 GAGGAGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1192
 DB 1319 GAGGAGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1378
 QY 1193 TGACAGATGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1252
 DB 1379 TGACAGATGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1438
 QY 1253 GTCGATGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1312
 DB 1439 ATGCTGATGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1498
 QY 1313 AGGAGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1372
 DB 1499 AGGAGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1558
 QY 1373 TATTCAGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1383
 DB 1559 TATTCAGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1618
 QY 1384 -----G 1384
 DB 1619 AAATGAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1678
 QY 1385 TCGTGAAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1444
 DB 1679 TCGTGAAGAGAGAGAGATTCGTCCTCTTCTGAGAGAGAGAGATTCGTCCTCTTCTGAG 1738

QY 1445 CTGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1504
 DB 1739 CCGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1798
 QY 1505 TTCTAAGTCAAGTAAAG 1564
 DB 1799 TTCTAAGTCAAGTAAAG 1858
 QY 1565 AAAAGTCAAGTCTTCAAG 1624
 DB 1859 AAAAGTCAAGTCTTCAAG 1918
 QY 1625 AG 1681
 DB 1919 AG 1978
 QY 1682 GAGTGAAG 1741
 DB 1979 GAGTGAAG 2038
 QY 1742 ACCCAACTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1801
 DB 2039 ATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 2089
 QY 1802 CCTATGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1823
 DB 2090 CACTGTGATTAATTCACGAA 2111

RESULT 10
 AAS86576
 ID AAS86576 standard; cDNA; 2332 BP.

AC AAS86576;
 DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #22380.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSB-) HYSBQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR MPI; 2001-639362/73.

XX P-PSDB; ABG22389.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

PS Claim 1; SEQ ID NO 22380; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostic as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pat_sequences

XX Sequence 2332 BP; 511 A; 621 C; 764 G; 436 T; 0 U; 0 Other;
 SQ
 Query Match 24.1%; Score 692; DB 5; Length 2332;
 Best Local Similarity 71.3%; Pred. No. 3.2e-205;
 Matches 1051; Conservative 0; Mismatches 335; Indels 88; Gaps 7;

QY 211 GCTTCACTTTTGGAGCAGTATGATGATTAAGGCACTGAGCAATGAGAGACCC 270
 DB |||||
 DB 370 GCTTCACTTTTACCGCTAGTATGAGCAAAAGGCGACCGAGCAATGAGAGACCT 429
 QY 271 AAGCTTCAACCCCAACGATGAAGATTCAAGAAACCTGGGGTTTCAAGAACCA 330
 DB |||||
 DB 430 AAGGCTCAACCCCAACGAGGAAGATTCAAGAAACCTGGGGTTTCAAGAACCA 489
 QY 331 ATTCCAAACGTGAGAGGTGAGAGACAGGAGCGGACCCCACTGAGCAGCAACCC 386
 DB |||||
 DB 490 ATCCCAAGCAGAGAGGCGGAGGAGCGGAGGAGCTGACCACTGAGACCCGACCCCA 549
 QY 387 --ACAGCAGATTAACCTCTCCCTGCGCGCAGTGAACGCAACCAACGATCTGAGAG 444
 DB |||||
 DB 550 CAGCAGCAGCTGAGGCTGTCCCTGCGGCGCAGTGGAGGACGACCCCAACGCACTGAGAG 609
 QY 445 GTAGAAGAGTTTCTTAACACGCTTCCGGGCGGAGGAGAAAGATGTCCTGTCCTG 504
 DB |||||
 DB 610 GTGAGCAGTTCTGACCATTTGCGGCGCGGAGGAGAGAGCATCTGTCTCTG 669
 QY 505 GAGGATTCAGTGAAGCCCACTCTTCAAGTCACTGATGATGAGAGCAGTTCGAGAGG 564
 DB |||||
 DB 670 GAGGATTCGATGAGCCCACTCTGCGCGGAGGAGGAGGAGGAGGAGGAGG 729
 QY 565 AGCTTGAAGCAGTTCGATGATGAGAGTGGCCCTGATCTGACTCTTA---GGGAA 621
 DB |||||
 DB 730 AGCTTGAAGCAGTTCGATGAGAGTGGCCCTGATCTGACTCTTA---GGGAA 789
 QY 622 GAACATCTGCTCTTCTGAAGAGGCAAAAGAGGTGAAGAGGAAAGACCTCTGAC 681
 DB |||||
 DB 790 GAACGACCGAGCTCTTCTGAAGAGGTAAGAGGAGGAGGAGGAGGAGGAGG 849
 QY 682 AGTGAAGTGAAGGCTCTTCAAGTGAAGGAACTTCAGAAACCGCTTCGAGAAAGGAGAG 741
 DB |||||
 DB 850 AGTGAAGGAGGCTCTTCAAGTGAAGGAACTTCAGAAACCGCTTCGAGAAAGGAGAG 909
 QY 742 CAAGAACCTTGAAGAGGCTCTTGAAGAGGCACTGAAATCCCTGAGAGAGAGGAGAG 801
 DB |||||
 DB 910 CAGGAGGCCCACTGAGAGGCGCCCTGAAGAGGATTCAGAGTCCGCTCCGAGAAAGAGCGCG 969
 QY 802 GAGGAAGTTTTCGCGGAACTGGAGAGTCCAAATAGCAGAGTCCGAGCAGAGAGAGAGACT 861
 DB |||||
 DB 970 GAGGAGGTTCCCGGAGACGTGTGGCTCCAGGSCCACTGACATCTGTGAGAGGAGCTCTG 1029
 QY 862 CTCTGTAAGCAGAGCCTGAGGCTAGTGAAGGAGCAAGTATCCAGTCAAGAGCAGATGAC 921
 DB |||||
 DB 1030 CCGAGTGAAGCAGAGCCCGAGAGACATCAAGGGGCTGTGTCTCCAGGCTGGGAAAGATGAC 1089
 QY 922 ATAGAAATCACTTGAAGAGGAGGAGGAGCACTCAAGGAAATACAGAGAGAAACCCAGGGA 980
 DB |||||
 DB 1090 AGAGAGAGTGAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1149

QY 981 AGCGGCAAAACCAAGCTGAGTGAAGTTTACGACCCCAATGCCCTGTACTGATCTG 1040
 DB |||||
 DB 1150 TTGGGGCCAGACCGAAGCTGAAATGAGAGGTTTACGACCCCAAGCGCTGTATGATTTG 1209
 QY 1041 CCGCAGCTTCAACAACAACAGTTTATGATCTGCTGTAATGAGTGAAGAGTGTTC 1100
 DB |||||
 DB 1210 CCGCAGCTTCAACAACAACAGTTTATGATTTGCTGTGACCGCTGTGAAGATGTGTCA 1269
 QY 1101 TGGTACCTGTGTGATTTCTGAGGCGCGAGGCGCTCCGAGAAAGAAAGGAGAG 1160
 DB |||||
 DB 1270 TGGCATTGTGTGCTATTTCTGAGGCTGAGAGGAGGCTTTTGAAGAGATGGGAGAG 1329
 QY 1161 CTACATCTGCCAAATTTGACCAATTTTCAAGTGAAGATGAGAGCAAAACGTAAGCGCCAC 1220
 DB |||||
 DB 1330 CTATATCTGCCAAATTTGACCAATTTCTGCAAGTGAAGATGAGATCTATTCAGAAACGCG 1389
 QY 1221 CAATGAGCAGACCTGTGGTCAAGTCTGTGGTGTCTGATGAGCAGACAC---TGACCA 1275
 DB |||||
 DB 1390 AGATCAGCAGAGAGCTTAAATGGAGACCTGGAGATCTTGAATGGACCAATTTTAC 1449
 QY 1276 AGCATAGGAGACGT--AGAGCAGAGTC---CGAGAAAGCAGAGGATTAAGGGTACG 1329
 DB |||||
 DB 1450 AGTATAGGAAACATTAGAGAGAGAGTCTTACCGAGAACCCAGGGGTTAAAGGTTAG 1509
 QY 1330 ATTGAGAGGAGCAAAACCCAGCGGCAAGAAACCTCAAGATATTCAGAGC----- 1382
 DB |||||
 DB 1510 ATTGAGAAAGCTGCAAAATCCAAAGTGGCAAGAAAGAACTCAAGATCTTCCAGCCTGAT 1569
 QY 1383 ----- 1382
 DB 1570 GAAATTCGTGTCTGATGAGACATCTGCTTCACACTTGTGACATGATGATGCA 1629
 QY 1383 TGTGTGAGAGCTCTGTGTGCTCTTAATGATGAGCCCTGGGTGTTCAAGTATGACCA 1442
 DB |||||
 DB 1630 GGTATAGAGGCGCTGTGTGCTCAAAATGATGTGGCCCGGTCTGTACGTGGCGCA 1689
 QY 1443 GCTGACTCTGTGTATTTGCAAGTAAATGATGATCTCAAAACGAGCAGTCAATCAG 1502
 DB |||||
 DB 1690 GCCGACCTCGGTGTATCTGCAATATGATGATCTCAAAACGCGGAGGAGCAATGAA 1749
 QY 1503 ATTCTAAGTTCAAGTAAAGAAACAAACCAAAACCCCAAGAGAAAGTCAAGACGAGCC 1562
 DB |||||
 DB 1750 GTTCTAAGCTCAAGTAAAGAAACCAAGCCAAAGCTTAAAGAAAGATGAGATGAGCC 1809
 QY 1563 AGAAAGTTCACTCTTCCAAATGCAAGTTCAG 1596
 DB |||||
 DB 1810 AGAGAGCCCAAGTCTTCCGAAATGCGGTCTCAG 1843

RESULT 11
 AAC77933
 ID AAC77933 standard; cDNA, 1764 BP.
 XX
 AC AAC77933;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human cancer associated gene sequence SEQ ID NO:327.
 XX
 XX Human: cancer associated gene; cancer antigen; detection; cancer;
 KM diagnosis; cytostatic; proliferative; vulnery; immunomodulator;
 KM antidiabetic; antiaesthetic; antirheumatic; antiarthritic; antiviral;
 KM antiinflammatory; antihypertoid; antiallergic; antibacterial; cardiac;
 KM dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
 KM vasectropic; antipneumatic; antiangiogenic; gene therapy; inflammation;
 KM immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KM allergic reaction; graft versus host disease; organ rejection;
 KM haemostatic; thrombolytic; cardiovascular disorder; infection;
 KM neurological disease; drug screening; ss.
 XX
 XX Homo sapiens.
 OS

OS Mus musculus.
 XX
 XX NC0200021991-A1.
 XX
 PD 20-APR-2000.
 XX
 PF 15-OCT-1999; 99WMO-US024206.
 XX
 PR 15-OCT-1998; 98US-0104436P.
 XX
 PA (GEMV) GENETICS INST INC.
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Bowman MR;
 DR WPI; 2000-317938/27.
 XX
 XX Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (ESTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders.
 XX
 PS Claim 1, Page 296-297, 803pp; English.
 XX
 CC AAA43426 to AAA45925 represent specifically claimed secreted expressed
 CC sequence tags (ESTs), isolated from human, mouse, chicken and rat tissue
 CC sources. The ESTs can have a range of activities depending on the
 CC tissues they were isolated from. The activities include: chemotactic;
 CC proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic;
 CC haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial;
 CC antifungal; antiviral; antidiabetic; antiaesthetic; vulnary; antitumor;
 CC osteoprotective; neuroprotective; nocotropic; antiparkinsonian; antipsoriatic;
 CC cerebroprotective; anticonvulsant; and antidepressant. The ESTs can be
 CC used for gene therapy and in vaccines. The ESTs are useful as probes for
 CC the identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the ESTs. Proteins encoded by the ESTs
 CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
 CC in the exemplification of the present invention
 XX
 XX
 SQ Sequence 408 BP; 116 A; 98 C; 122 G; 72 T; 0 U; 0 Other;
 Query Match 13.4%; Score 382.8; DB 3; Length 408;
 Best Local Similarity 99.5%; Pred. No. 9.3e-109;
 Matches 384; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 197 CCAAGGTTTCAGCTTTCAGTTTGGACAGTATGATGATTAAGGGCAGCTGACCA 256
 DB 19 CTAAGGTTTCAGCTTTCAGTTTGGACAGTATGATGATTAAGGGCAGCTGACCA 78
 QY 257 ATGAGGAAGCCCAAGGCTATCAACCCACAGTAAGAGTTCAGGAAACCTGGGGTT 316
 DB 79 ATGAGGAAGCCCAAGGCTATCAACCCACAGTAAGAGTTCAGGAAACCTGGGGTT 138
 QY 317 TTCCAGAAACCAAGCATTCCAAACCTGAGGGTGCAGAGACACGAGGCGGACCCAGTG 376
 DB 139 TTCCAGAAACCAAGCATTCCAAACCTGAGGGTGCAGAGACACGAGGCGGACCCAGTG 198
 QY 377 AGCAGCAACCAAGCATTTCCTCCCTGCGCGCAGTGGACGCGCAACCAAAAGCTA 436
 DB 199 AGCAGCAACCAAGCATTTCCTCCCTGCGCGCAGTGGACGCGCAACCAAAAGCTA 258
 QY 437 CTGAGAGGGTGAAGAGTTTCTTACCAAGGTTGCGGCGCAGAGGAAAAAGATGCGCG 496
 DB 259 CTGAGAGGGTGAAGAGTTTCTTACCAAGGTTGCGGCGCAGAGGAAAAAGATGCGCG 318
 QY 497 TGTCCCTGAGAGATTCCAGTAGGCCCAATTTCCAGTCACTGATGTGAGACAGCTT 556

DB 319 TGTCCCTGAGAGATTCCAGTAGGCCCAATTTCCAGTCACTGATGTGAGACAGCTT 378
 QY 557 CCGAGGGAGGCGTTGAAGCAGTTCT 582
 DB 379 CCGAGGGAGGCGTTGAAGCAGTTCT 404
 RESULT 13
 ID AAS86575 standard; cDNA; 2986 BP.
 XX
 AC AAS86575;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #22379.
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 XX
 XX WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PP 30-MAR-2001; 2001WO-US008631.
 XX
 XX 31-MAR-2000; 2000US-00540217.
 XX
 PR 23-AUG-2000; 2000US-00649167.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 XX WPI; 2001-639362/73.
 DR P-PSDB; ABG22388.
 XX
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensic, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 1; SEQ ID NO 22379; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensic, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2986 BP; 798 A; 796 C; 781 G; 611 T; 0 U; 0 Other;
 Query Match 13.1%; Score 376; DB 5; Length 2986;
 Best Local Similarity 68.6%; Pred. No. 4.9e-106;
 Matches 681; Conservative 0; Mismatches 190; Indels 121; Gaps 7;

```
QY 924 AGAAAAATCATGTTGAGAGGAGGCGACTCAGAGGAAATACAGAGAAAACCCAGGAGAC 983
DB 2 AGAGAGTAGTTGAGAGGAGGAGGCGCTCAGACATCAAGATGAGAGAGCTTGAGACTT 61
QY 984 GGGCAAAACCAAGCCCTGAGTGTGAGTTTACGACCCCAATGCCCTGTACTGTGACTGTGCG 1043
DB 62 GGGCCGACCCGAAAGCTGAAATGTAGGGTTTACGACCCCAAGCCCTGTATTTGCAATTTGCG 121
QY 1044 CCAGCTCTCAACAACAAGGTTTATGATCTGCTGTGATCGGTGTGAGAGAGTGTTCATG 1103
DB 122 CAGGCTCTCAACAACAAGGTTTATGATCTGCTGTGATCGGTGTGAGAGAGTGTTCATG 181
QY 1104 TGACTGTGTGGTATTTTGAAGGCGGAGCGGCTCTGTGAAAGGAAACGGGAGAGACTA 1163
DB 182 CGATTGTGTGGCATTTTGAAGGCTGAGAGGAGGCTTTTGAAAGGAAATGGGAGAGACTA 241
QY 1164 CATCTGCCCAATTTGACATTTTGAAGTGTGAGAGTGTGAGCAACGGTAGCCCAACAA 1223
DB 242 TATCTGCCCAATTTGACATTTTGAAGTGTGAGAGTGTGAGCAACGGTAGCCCAACAA 301
QY 1224 TGAGCAGAGACTCTGGGTGAGATCTGTGGGTGCTGATGGCAGACTGCAACAGCATAG 1283
DB 302 TCAGCAGAGAGCTAAATGAGAGCTTGAGATGCTGATGGCAGACTGCAACAGCATAG 361
QY 1284 GACAGTAGAGCAGAAAGTCCGAGAGAGACAGAGGCAATAAAGGTTAGATTTGAGAGGCA 1343
DB 362 AACAAATAGAGCAGAAAGTCTAGAGAGACCAAGGAGATTAAGGTTAGAGAAAGCTG 421
QY 1344 AAACCCAGCGGCAAGAAAAATCTCAAGATTTCCAGCCT----- 1383
DB 422 AAATCCAAATGTGCAAGAAAGAACTCAAGATCTTCCAGCCTGCTCCGGGCTGTCCCA 481
QY 1384 ----- 1383
DB 482 CCAAGCTGCTGTCTCTGCGAGGATTTGAAATTCGTGTGTCTAGAGAGCATCTTCCTT 541
QY 1384 -----GTCTGAGAGGCTCTCTGTGCTCTTCAATGAT 1415
DB 542 CACAGCTTTGCACTGATTAAGCTGCAAGAGGATTAAGAGGCGCTGTGCTCAAAATGTAT 601
QY 1416 TGG-CCCTGGGTGTTCCAGTGTGACAG-CTGATCTGTGTATTTGCA-GTATGTACTG 1472
DB 602 TGGCCCGGGGTGCTGTCACTGTGCGCAGCCCGACTCGGTGTACTGTGAGGTATGACTG 661
QY 1473 CATTTCTAAACGCGAGC-AGCTACATGAGATTTCTAAGTTCAAGTAAAGAACAAAAA 1531
DB 662 TATCTCAAAACGCGCAGAGGACATTAAGTTTCTAAGCTCAAGTAAAGAAAGAGAG 721
QY 1532 CAAAAACCAAGAGAAAGTCAAGACAGAGCCAGAAAGTTCAGTCTTCCAAATGTGAGTG 1591
DB 722 CAAAGCTTAAGAGAAAGATGAGATGAGCCAGAGAAAGCCAGTCTTCCGAAATGTGAGTG 781
QY 1592 TTCAAGTGGGATTAATCTTTCTGTGCACAAGAGATGAGGTGAGAGAGAGAGAGAA 1651
DB 782 CTCAGGCGAGTATTAAATCTCTTCTGTGCACAAGAGACAGAGTCTCCAGAAAAAGAG 841
QY 1652 ACCCAGTAAAGAA---AGTATGCTGTGCTTCCAGAGTGAAGCTTTGAGAGAGAGAG 1708
DB 842 CCAAGTAGAAGAGAGAGTGTGTCTTCCGCGAGTGAAGAGCTGGAGAGAGAGAGAG 901
QY 1709 CCTGTGAGAGCAGACACCATCTGGGCAAGTGAACCAACTACATGTCTGTGAAGCCAG 1768
DB 902 CTTGTGAGAGCAGACACCATCTGGGCAAGTGAACCAACTACATGTCTGTGAAGCCAG 955
QY 1769 AGAAGCCAGAGAGAGCCACTGTGAGCC 1800
DB 956 TAAAGCCAGAGAGAGAGCTGTGCTCCCTGCGCC 987
```

RESULT 14
AAS86574
ID AAS86574 standard; cDNA; 503 BP.

```
XX AAS86574;
AC 13-FEB-2002 (first entry)
DT
DB DNA encoding novel human diagnostic protein #22378.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001MO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Dmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX P-PSDB; ABG22387.
XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostic, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX
XX Claim 1, SEQ ID NO 22378; 103bp; English.
XX
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantifying a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX
XX Sequence 503 BP; 165 A; 107 C; 138 G; 93 T; 0 U; 0 Other;
XX
XX
XX Query Match 12.1%; Score 347.8; DB 5; Length 503;
XX Best Local Similarity 80.7%; Pred. No. 1e-97;
XX Matches 406; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
XX
XX
XX 1132 GGGCGGCTCTGTGAAAGGAGAGGAGGAGGAGTACATGCGCCAAATTTGACACCATTTTGCA 1191
DB 1 GGGAGGCTTTTGAAGAGATGGAGAGGAGTATATCTGCCAAATCTGACACATTTTGCA 60
QY 1192 GTGCAAGATGAGCAAAACGTTAGCGCCACCAATGAGCAGAGCTTGGGTGAGATGTGTG 1251
DB 61 GTGCAAGATGAGCAATTCATTCAGAAAGCGCAGATCACAGAGACTTAAATGTGAGACCTGGA 120
QY 1252 GGTGCTGATGAGCAGACTGCAACAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1311
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2004, 04:42:12 ; Search time 10916.3 Seconds
(without alignment) 11383.416 Million cell updates/sec

Title: US-09-787-016a-2

Perfect score: 2867
Sequence: 1 tccgtgtagctcgtcgaat.....tttgttctcttgcgtcggg 2867

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 2167516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: GenBank:
- 2: gb_ba:
- 3: gb_hg:
- 4: gb_in:
- 5: gb_om:
- 6: gb_ov:
- 7: gb_pat:
- 8: gb_ph:
- 9: gb_pl:
- 10: gb_pr:
- 11: gb_ro:
- 12: gb_ry:
- 13: gb_un:
- 14: gb_vl:
- 15: gb_da:
- 16: gb_fun:
- 17: gb_hum:
- 18: gb_in:
- 19: gb_mu:
- 20: gb_om:
- 21: gb_or:
- 22: gb_ov:
- 23: gb_pat:
- 24: gb_ph:
- 25: gb_pl:
- 26: gb_ro:
- 27: gb_ry:
- 28: gb_un:
- 29: gb_vl:
- 30: gb_da:
- 31: gb_fun:
- 32: gb_hum:
- 33: gb_in:
- 34: gb_mu:
- 35: gb_om:
- 36: gb_ov:
- 37: gb_pat:
- 38: gb_ph:
- 39: gb_pl:
- 40: gb_ro:
- 41: gb_ry:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2867	100.0	2867	6	BD244713	BD244713 Genes enc
2	2867	100.0	2867	6	AX023370	AX023370 Sequence
3	2602	90.8	2602	10	MMU218332	MMU218332 Mus muscu
4	1770.4	61.8	4906	10	AK129117	AK129117 Mus muscu
5	1241.2	43.3	10463	10	AL732560	AL732560 Mouse DNA
6	1023.6	35.7	2610	6	BD244712	BD244712 Genes enc
7	1023.6	35.7	2610	6	AX023369	AX023369 Sequence
8	1013.6	35.4	2725	9	BC014489	BC014489 Homo sapi
9	1013.6	35.4	2725	9	BC014489	BC014489 Homo sapi
10	1013.6	35.3	2772	6	BD157094	BD157094 Primer to
11	1013.6	35.3	2772	6	BD157094	BD157094 Primer to
12	1013.6	35.3	2772	6	BD157094	BD157094 Primer to
13	985.4	29.8	221547	2	AC108337	AC108337 Rattus no
14	853.4	26.1	272404	2	AC108337	AC108337 Rattus no
15	748.2	24.5	69252	2	AC101519	AC101519 Mus muscu
16	703.4	24.5	69252	2	AC101519	AC101519 Mus muscu
17	449.8	15.7	6692	9	AB002331	AB002331 Human DNA
18	365.4	12.7	28519	9	HSJ563814	HSJ563814 Mus muscu
19	324.4	11.3	69252	2	AC101519	AC101519 Mus muscu
20	230.4	8.0	160241	9	HS88517	HS88517 Xenopus l
21	230.2	8.0	2428	5	BC060442	BC060442 Sequence
22	209.8	7.3	681	6	BD149997	BD149997 Primer to
23	209.8	7.3	681	6	BD149997	BD149997 Primer to
24	184.6	6.4	812	11	BV013961	BV013961 Danio rer
25	108.6	3.8	198133	2	BX296530	BX296530 Sequence
26	77.8	2.7	7218	6	I66494	I66494 Sequence
27	70.2	2.4	61204	2	AC017132	AC017132 Drosophi
28	70.2	2.4	61204	2	AC017132	AC017132 Drosophi
29	70.2	2.4	225655	3	AB003695	AB003695 Drosophi
30	69.6	2.4	2161	3	AK114522	AK114522 Clona int
31	57.8	2.0	2308	9	BC015733	BC015733 Homo sapi
32	57.8	2.0	2308	9	BC015733	BC015733 Homo sapi
33	57.8	2.0	2320	6	AX086854	AX086854 Sequence
34	57.8	2.0	2320	6	AX086854	AX086854 Sequence
35	57.8	2.0	2387	10	BC030938	BC030938 Homo sapi
36	57.8	2.0	2487	9	AF149758	AF149758 Homo sapi
37	57.8	2.0	2487	9	AB031069	AB031069 Homo sapi
38	57.8	2.0	2509	10	AY099096	AY099096 Mus muscu
39	57.8	2.0	2967	9	HS8807333	HS8807333 Homo sapi
40	56.8	2.0	1310	3	AY089443	AY089443 Homo sapi
41	56	2.0	10732	6	E32986	E32986 Gene encodi
42	55.4	1.9	2868	9	AK055678	AK055678 Homo sapi
43	55.4	1.9	5975	6	AX329603	AX329603 Sequence
44	55.4	1.9	6256	6	AR338778	AR338778 Sequence
45	55.4	1.9	6936	9	D87685	D87685 Human mRNA

ALIGNMENTS

RESULT 1
BD244713
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

BD244713 2867 bp DNA linear PAT 17-JUL-2003
Genes encoding for the human and murine death
inducer-obliator-1.
BD244713 GI:33054483
BD244713.1 GI:33054483
JP 2002526040-A/2.
Mus sp.
Mus sp.
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2867)
Alonso, C.M., Domingo, D.G., Grandien, A., Leonardo, B. and Martinez, P.
Genes encoding for the human and murine death inducer-obliator-1

JOURNAL Patent: JP 2002526040-A 2 20-AUG-2002;
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS
OS Mus sp. (murine)

Patent: JP 0002526040-A 2 20-AUG-2002;
 CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS
 OS Mus sp. (murine)
 PN JP 2002526040-A/2
 PD 20-AUG-2002
 PF 10-SEP-1999 JP 2005570314
 PR 10-SEP-1998 SE 9803069-5 17-SEP-1998 US 60/100873 P1
 CARLOS MARTINEZ ALONSO, DAVID GARCIA DOMINGO, ALF GRANDIEN, P1
 ESTHER LEONARDO,
 PEDRO MARTINEZ
 PC C12M15/09, A61K38/00, A61K45/00, A61K48/00, A61P3/10, A61P7/00, A61P23/00,
 PC C12N5/00, A61P37/06, A61P43/00, C07K14/47, C07K16/18, C12N1/21, PC C12N5/10,
 PC A61P35/00, A61P37/06, A61P43/00, C07K14/47, C07K16/18, C12N1/21, PC C12P21/02, C12Q1/02, C12Q1/66, G01N33/15, G01N33/50/ C12P21/08, PC C12P21/02, C12Q1/02, C12Q1/66, G01N33/15, G01N33/50/ C12P21/08, PC

PC	C12N5/00,A61K37/02	Genes encoding for the human and murine death inducer-CC
CC		
PH	Key	Location/Qualifiers
PT	source	1.2867
PT		/organism='Mus sp. (murine)'.

FEATURES	Location/Qualifiers
source	1..2867
	/organism="Mub sp."
	/mol_type="genomic DNA"
	/db_xref="taxon:10095"

ORIGIN

Query March	100.0%	Score 2867;	DB 6;	Length 2867;
Best Local	100.0%	Pred. No. 0;		
Matches 2867;	0;	Mismatches	0;	Gaps 0;

QY	I	TCCTGTGTAGCTCTGTGAAATATGGCTGCGGGATTCGCGCGCGCGGGAGACTTGTTTAAGAG	60
Db	1	TCCCTGTGTAGCTCTGTGAAATATGGCTGCGGGATTCGCGCGCGCGGGAGACTTGTTTAAGAG	60
QY	61	CAGTCCCCACTGTCCCTTTGTGTGTGTGAAGCTCGGAAATCTTCAATGTAAATGAACGTG	120
Db	61	CAGTCCCCACTGTCTCCCTTTGTGTGTGTGAAGCTCGGAAATCTTCAATGTAAATGAACGTG	120
QY	121	ACTGACAAACAGTGGGGGTAGGCTTGACCGTGTGCTTGAACCTGCCCCAGGCTTAAATTT	180
Db	121	ACTGACAAACAGTGGGGGTAGGCTTGACCGTGTGCTTGAACCTGCCCCAGGCTTAAATTT	180
QY	181	TATGTAGATTTCCAGCCAAAGGTTCCAGCTTCAAGTTTGTGGACAGGTATGATGAT	240
Db	181	TATGTAGATTTCCAGCCAAAGGTTCCAGCTTCAAGTTTGTGGACAGGTATGATGAT	240
QY	241	AAAGGGCACTGTAGCAATGAGAAAGACCCAGGCTATCCAAACCAACAGTAAAGAGTTT	300
Db	241	AAAGGGCACTGTAGCAATGAGAAAGACCCAGGCTATCCAAACCAACAGTAAAGAGTTT	300
QY	301	AGGAAAACCTGTGGGTTTTGGAAGAACACGATTCGCAACGTGAGGGGTGCAGAGACAG	360
Db	301	AGGAAAACCTGTGGGTTTTGGAAGAACACGATTCGCAACGTGAGGGGTGCAGAGACAG	360
QY	361	GAGCGGAGCCCACTGAGACAGACACACAGCAGCATTAACCTTCTCTGCGCGCAGTGA	420
Db	361	GAGCGGAGCCCACTGAGACAGACACACAGCAGCATTAACCTTCTCTGCGCGCAGTGA	420
QY	421	CGGCAACCAAAACGTACTGAGAGGGTGAAGAAGTTCTTACAGAGGTTGCGCGCGAGAG	480
Db	421	CGGCAACCAAAACGTACTGAGAGGGTGAAGAAGTTCTTACAGAGGTTGCGCGCGAGAG	480
QY	481	AAAAAGATGTGCGGTGTCTCCCTGAGAGTTCCAGTGAAGCCACACTTTCACAGTCACT	540
Db	481	AAAAAGATGTGCGGTGTCTCCCTGAGAGTTCCAGTGAAGCCACACTTTCACAGTCACT	540
QY	541	GATGTGAGAGACAGTTTCCAGGGGAGAGCTTGAAGCAGTTCTGAGATTCAGAAATGTGGCCCT	600
Db	541	GATGTGAGAGACAGTTTCCAGGGGAGAGCTTGAAGCAGTTCTGAGATTCAGAAATGTGGCCCT	600

QY	601	GTATCTGATCTCTTAAGGGAAAGAAACATCTCTCTCTTCTGAAAAGGGCAAAAGAGAGTGA	660
Db	601	GTATCTGATCTCTTAAGGGAAAGAAACATCTCTCTCTTCTGAAAAGGGCAAAAGAGTGA	660
QY	661	GAGGAGAGAGCACTCTGACAGTGAACAGTAGTGGCTTACGTTGAGGAACCTTCAGAAC	720
Db	661	GAGGAGAGAGCACTCTGACAGTGAACAGTGAATGGCTTACGTTGAGGAACCTTCAGAAC	720
QY	721	CGCCTTCGGAAGAAAGCGGAGAGCAAGAACTCTGTGAGAGAGTCCCTGAGAGGGAGTCAAGAT	780
Db	721	CGCCTTCGGAAGAAAGCGGAGAGCAAGAACCTGTGTGAGAGAGTCCCTGAGAGGGAGTCAAGAT	780
QY	781	CGCCTGAGGAAGAAAGCGGAGAGCAAGAACTCTGCCGAACTGAGGAGTGTCAATAAGGC	840
Db	781	CGCCTGAGGAAGAAAGCGGAGAGCAAGAACTCTGCCGAAACTGAGGAGTGTCAATAAGGC	840
QY	841	AGTCCCGAGCAGACAGACTCTCTCTGTAGCAGAGGCTGAGGCTAGTCAGGGACCACTG	900
Db	841	AGTCCCGAGCAGACAGACTCTCTCTGTAGCAGAGGCTGAGGCTAGTCAGGGACCACTG	900
QY	901	TCCAGTCAAGAGACAGATGACATGAGAAATCAGTTGGAGGGAAGGCCACTCAGGGAAAT	960
Db	901	TCCAGTCAAGAGACAGATGACATGAGAAATCAGTTGGAGGGAAGGCCACTCAGGGAAAT	960
QY	961	ACAGAGGAAAAACCCAGGGAAAGCGGGCAAAACCAAGCCTGAGTGAAGTTTACGACCC	1020
Db	961	ACAGAGGAAAAACCCAGGGAAAGCGGGCAAAACCAAGCCTGAGTGAAGTTTACGACCC	1020
QY	1021	AATGCCCTCTGTATCTGCATCTGCCGACGCTTCAACAACAAGGTTTATGATCTGCTGTAT	1080
Db	1021	AATGCCCTCTGTATCTGCATCTGCCGACGCTTCAACAACAAGGTTTATGATCTGCTGTAT	1080
QY	1081	CGGTGTGAGAGTGGTTCATGATGTGACTGTGTGGGTATTTCTGAGGCCCAGAGCGGCTC	1140
Db	1081	CGGTGTGAGAGTGGTTCATGATGTGACTGTGTGGGTATTTCTGAGGCCCAGAGCGGCTC	1140
QY	1141	CTGGAAGGAAACGGGGAAAGACTACATCTGCCCAATTGCAACATTTTGCAGATGCAAGAT	1200
Db	1141	CTGGAAGGAAACGGGGAAAGACTACATCTGCCCAATTGCAACATTTTGCAGATGCAAGAT	1200
QY	1201	GAGCAAAACGGTACGGCAACCAATGAGACAGGACTCTGGGTCAATCTGTGGGTCTGAT	1260
Db	1201	GAGCAAAACGGTACGGCAACCAATGAGACAGGACTCTGTGGGTCAATCTGTGGGTCTGAT	1260
QY	1261	GGCAACAGCTGCACAAACATAGGACAGTAGAGCAGAAAGTCCGGAGAAAGCAAGGGCAT	1320
Db	1261	GGCAACAGCTGCACAAACATAGGACAGTAGAGCAGAAAGTCCGGAGAAAGCAAGGGCAT	1320
QY	1321	AAGGTATGATATGGAAGAGCAGCAACCCCAACGGCAAGAAAACCTCAAGATATTCAG	1380
Db	1321	AAGGTATGATATGGAAGAGCAGCAACCCCAACGGCAAGAAAACCTCAAGATATTCAG	1380
QY	1381	CCTGTCTGAGAGCTCTCTGTGTCTCTTAAATGCAATTGGCCCTGTGGTTCACAGTGAACA	1440
Db	1381	CCTGTCTGAGAGCTCTCTGTGTCTCTTAAATGCAATTGGCCCTGTGGTTCACAGTGAACA	1440
QY	1441	CAGCTTGACTCTGTGTATTTGCAATGATATGACTGCATTTCTCAAAACACGACAGCACTATG	1500
Db	1441	CAGCTTGACTCTGTGTATTTGCAATGATATGACTGCATTTCTCAAAACACGACAGCACTATG	1500
QY	1501	AGATTTTCAAGTTCAAGGTAAAGAACAAAAAACAAACCCCAAGGAAAGGTCAACAGAAC	1560
Db	1501	AGATTTTCAAGTTCAAGGTAAAGAACAAAAAACAAACCCCAAGGAAAGGTCAACAGAAC	1560
QY	1561	CCAGAAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAATCTCTTCTGTG	1620
Db	1561	CCAGAAAAAGTTCAGTCTTCCAAAATGCAGTGTTCAGGTGGGGATTAAATCTCTTCTGTG	1620
QY	1621	CACAAAGAGACTAGGCTCAGAGAAAAGGGAAAAACCACTGAGAAAGATGTCTGGCTTCC	1680
Db	1621	CACAAAGAGACTAGGCTCAGAGAAAAGGGAAAAACCACTGAGAAAGATGTCTGGCTTCC	1680
QY	1681	AGGAGTGAACCTTCGGAAAGAGAGCGCTGTGAGAGCAACCATCTCTGGCAAGT	1740

Db	1661	AGAGGTGAGACTTCTGGAGAAAGAGCACCCTGTGAGACACACACACACTCTCGGACAGT	1740
Oy	1741	GACCAACAATCAATGCTGTGAAGCAGAGAAAGCCAGAGAGCCCACTGCATCTCGCCC	1800
Db	1741	GACCAACAATCAATGCTGTGAAGCAGAGAAAGCCAGAGAGCCCACTGCATCTCGCCC	1800
Oy	1801	ACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGCTTCCAGGCGCTTCCCATCAT	1860
Db	1801	ACCCTATTGAGTAAATGTACGTATCACCCAAAGGCTGCTTCCAGGCGCTTCCCATCAT	1860
Oy	1861	CTGGGTGGCTGCTCGGGGCTGTCTTAGGACACAGAGCTCGGGGTGTCTGGGTCTATGTGA	1920
Db	1861	CTGGGTGGCTGCTCGGGGCTGTCTTAGGACACAGAGCTCGGGGTGTCTGGGTCTATGTGA	1920
Oy	1921	GCACGACATCACTGCACGACGAGAAAGAGATACCAAGATGCTCTGACCCCAAGTTC	1980
Db	1921	GCACGACATCACTGCACGACGAGAAAGATACCAAGATGCTCTGACCCCAAGTTC	1980
Oy	1981	CTGCCTAGCCCTGTGGAGGCTCTCTGGGTGGTTCCTAAAGACCTGTGTAGGCTCATGTTG	2040
Db	1981	CTGCCTAGCCCTGTGGAGGCTCTCTGGGTGGTTCCTAAAGACCTGTGTAGGCTCATGTTG	2040
Oy	2041	GAGGCAATTTCTTATTTCAGTTTTAGGCTTGGTAAATTTGAATGCTCTCAACAAGA	2100
Db	2041	GAGGCAATTTCTTATTTCAGTTTTAGGCTTGGTAAATTTGAATGCTCTCAACAAGA	2100
Oy	2101	TGGGACACATCAGTGCATGTCAGACTCAGATTCAAGCTTGAGTTCCTCACAGAGACAGC	2160
Db	2101	TGGGACACATCAGTGCATGTCAGACTCAGTTCAAGCTTGAGTTCCTCACAGAGACAGC	2160
Oy	2161	CCAGCAACAACATGAGATTTTCCACGTTACTAATTTTCTTCAACAATGTCACACTG	2220
Db	2161	CCAGCAACAACATGAGATTTTCCACGTTACTAATTTTCTTCAACAATGTCACACTG	2220
Oy	2221	CTCTTCTATAGAAAATGATTTTTTTTCAATGATTAGGCAATTTTGAAGAAGACAGTGA	2280
Db	2221	CTCTTCTATAGAAAATGATTTTTTTTCAATGATTAGGCAATTTTGAAGAAGACAGTGA	2280
Oy	2281	CAATTGGTGTATTTTGCATATCAACACCTTACATCAATCAAGGTTTGAATTTTGAAGGTC	2340
Db	2281	CAATTGGTGTATTTTGCATATCAACACCTTACATCAATCAAGGTTTGAATTTTGAAGGTC	2340
Oy	2341	AATGCTGCTCAGATCCCATGGTGAATGTCGTGACCAATACACTACAGCATATTC	2400
Db	2341	AATGCTGCTCAGATCCCATGGTGAATGTCGTGACCAATACACTACAGCATATTC	2400
Oy	2401	TCATCAAAAGTGAAATGTGGGTGTAGCTTCCAGACTCGTAGATGCCCAAGATTTCAA	2460
Db	2401	TCATCAAAAGTGAAATGTGGGTGTAGCTTCCAGACTCGTAGATGCCCAAGATTTCAA	2460
Oy	2461	TCAGTGTATTATGTTTGAATAACTTTTGAACCCAGATTAAGTCACTAGATTTGAGC	2520
Db	2461	TCAGTGTATTATGTTTGAATAACTTTTGAACCCAGATTAAGTCACTAGATTTGAGC	2520
Oy	2521	CATTCTGTTCATCTGCATGAACAAGTACACTTTTCTTACTTTTGTATGACGTGGTTTCA	2580
Db	2521	CATTCTGTTCATCTGCATGAACAAGTACACTTTTCTTACTTTTGTATGACGTGGTTTCA	2580
Oy	2581	GTGAATTTTCACTGTGTGTTCTCATGATAGTGTGTTGTCATCTGTGCCACCTCAGCC	2640
Db	2581	GTGAATTTTCACTGTGTGTTCTCATGATAGTGTGTTGTCATCTGTGCCACCTCAGCC	2640
Oy	2641	CCAAGTAAAGGGGTCTCATGCTGACCTTAAAGCTGCGCTTTAAATTTTCTGTAAAGTTGGGGTTC	2700
Db	2641	CCAAGTAAAGGGGTCTCATGCTGACCTTAAAGCTGCGCTTTAAATTTTCTGTAAAGTTGGGGTTC	2700
Oy	2701	TCCTGCCCCCCTGTAAAGAGTGTGTAACCTTACAGGACACACACTGTATATGCTGTG	2760
Db	2701	TCCTGCCCCCCTGTAAAGAGTGTGTAACCTTACAGGACACACACTGTATATGCTGTG	2760
Oy	2761	GGGTGTGAACCTCAGAGTGTGGAACACAGACTCTACAGCTAGCTGACGTTCTAGCCATG	2820

Db	2761	GGGGTGAACCTCAGAGTGTGGACACAGAGCTACACAGCTGAGCTGACGTTCTAGCCATG	28252
QY	2821	GTAACCTGAAGCTGCTGATCTTCTGTCCTTTGTTCTTTGTCGCGG	2867
Db	2821	GTAACCTGAAGCTGCTGATCTTCTGTCCTTTGTTCTTTGTCGCGG	2867
RESULT 2			
AX023370			
LOCUS	AX023370	2867 bp	DNA linear PAT 15-SEP-2000
DEFINITION	Sequence 2 from Patent WO0015787.		
ACCESSION	AX023370		
VERSION	AX023370.1	GI:10183782	
KEYWORDS			
SOURCE	Mus sp.		
ORGANISM	Mus sp.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Leonardo, B., Martinez, P., Alonso, C. M., Domingo, D. G. and Grandien, A.		
TITLE	Genes encoding for the human and murine death in dicer-oblierator-1		
JOURNAL	Patent: WO 0015787-A 2 23-MAR-2000; LEONARDO ESTHER (ES); MARTINEZ PEDRO (ES); CONSELJO SUPERIOR INVESTIGACION (ES); DOMINGO DAVID GARCIA (ES); MARTINEZ ALONSO CARLOS (ES); BANNERMAN DAVID GARDNER (GB); GRANDIEN ALF (SE)		
FEATURES	Location/Qualifiers		
source	1..2867		
ORIGIN	/organism="Mus sp." /mol_type="unassigned DNA" /db_xref="taxon:10095"		
Query Match	100.0%; Score 2867; DB 6; Length 2867;		
Best Local Similarity	100.0%; Pred. No. 0;		
Matches 2867; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	TCGCTGTGAGCTCTGGAATGCTGCGGGATCCCGCGCGCGGGAGCTTGTAAAGG	60
Db	1	TCGCTGTGAGCTCTGGAATGCTGCGGGATCCCGCGCGCGGGAGCTTGTAAAGG	60
QY	61	CAGTCCCACTGTGCCCTTGTGTTGTAAGCTCCGGAATCTTCTCATATGAACTGTG	120
Db	61	CAGTCCCACTGTGCCCTTGTGTTGTAAGCTCCGGAATCTTCTCATATGAACTGTG	120
QY	121	ACTGACAAAGTGGGGTGAAGGCTTGGCGCTGCTGACCTGCGCCAGGCTATTAATT	180
Db	121	ACTGACAAAGTGGGGTGAAGGCTTGGCGCTGCTGACCTGCGCCAGGCTATTAATT	180
QY	181	TATGTAGGATTTCCAGCCAAAGTTTCCAGCTTTCAGTTTGTGGACAGTATGATGAT	240
Db	181	TATGTAGGATTTCCAGCCAAAGTTTCCAGCTTTCAGTTTGTGGACAGTATGATGAT	240
QY	241	AAAGGGCACTGAGCAATGAGGAAGCACCCAGGCTATCAAAACCACAGTAGAGATTG	300
Db	241	AAAGGGCACTGAGCAATGAGGAAGCACCCAGGCTATCAAAACCACAGTAGAGATTG	300
QY	301	AGGAAAACCTGGGTTTCCAGAAACCAAGATTGGCCAAAGTGAAGGTGACAGAGACG	360
Db	301	AGGAAAACCTGGGTTTCCAGAAACCAAGATTGGCCAAAGTGAAGGTGACAGAGACG	360
QY	361	GAGCGGACCCCAAGTGAAGCAACACAGAGATTAACCTTCCCTGCGCGGAGTGA	420
Db	361	GAGCGGACCCCAAGTGAAGCAACACAGAGATTAACCTTCCCTGCGCGGAGTGA	420
QY	421	CGGCAACCAAAAGCTACTGAGAGGGTGAAGAGTTTCTTACACAGTTTGGCGCGG	480
Db	421	CGGCAACCAAAAGCTACTGAGAGGGTGAAGAGTTTCTTACACAGTTTGGCGCGG	480
QY	481	AAAAAGAAATGTCGCGGTGTCCTGAGAGATTCACATGAGGCGGCATCTTCCACAGTCA	540
Db	481	AAAAAGAAATGTCGCGGTGTCCTGAGAGATTCACATGAGGCGGCATCTTCCACAGTCA	540

QY 541 GATGTGAGACAGCTTCCGAGGGAGGCTTGAAAGCAGTTCTGAGATCAGAGTGGCCCT 600
DB 541 GATGTGAGACAGCTTCCGAGGGAGGCTTGAAAGCAGTTCTGAGATCAGAGTGGCCCT 600
QY 601 GTATCTGACTCTTATGAGGAAAGAAATCTCTGCTCTTCTGAAAGGCAAAAGAGTGA 660
DB 601 GTATCTGACTCTTATGAGGAAAGAAATCTCTGCTCTTCTGAAAGGCAAAAGAGTGA 660
QY 661 GAGGAAGAGACACCTCTGACATGACATGATGAGCTTACGTTGAGAGAACTTCAAGAC 720
DB 661 GAGGAAGAGACACCTCTGACATGACATGATGAGCTTACGTTGAGAGAACTTCAAGAC 720
QY 721 CGCTTTCGGAAGAAAGCGAGACAGAACTTCTGAGAGAGTCCCTGAGAGGACATGAGAT 780
DB 721 CGCTTTCGGAAGAAAGCGAGACAGAACTTCTGAGAGAGTCCCTGAGAGGACATGAGAT 780
QY 781 CGCTTTCGGAAGAAAGCGAGACAGAACTTCTGAGAGAGTCCCTGAGAGGACATGAGAT 840
DB 781 CGCTTTCGGAAGAAAGCGAGACAGAACTTCTGAGAGAGTCCCTGAGAGGACATGAGAT 840
QY 841 AGTCCGAGACAGACAGCTCTCTGTAAGCAGAGCTTGAAGGCTGAGGCTGAGGACAGTG 900
DB 841 AGTCCGAGACAGACAGCTCTCTGTAAGCAGAGCTTGAAGGCTGAGGCTGAGGACAGTG 900
QY 901 TCCCACTGAGAGACAGATGACATGAGAAATCAGTTGAAAGGAGAGGACATCAGGAAAT 960
DB 901 TCCCACTGAGAGACAGATGACATGAGAAATCAGTTGAAAGGAGAGGACATCAGGAAAT 960
QY 961 ACAGAGGAAACCCCGAGGAGAGCGGCAAAACCAAGCTGAGTGTGAGGTTTACAGCC 1020
DB 961 ACAGAGGAAACCCCGAGGAGAGCGGCAAAACCAAGCTGAGTGTGAGGTTTACAGCC 1020
QY 1021 AATGCTCTGTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
DB 1021 AATGCTCTGTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
QY 1081 CGGTGTGAGAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
DB 1081 CGGTGTGAGAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
QY 1141 CTGGAAGGAAAGCGGGAAGACTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
DB 1141 CTGGAAGGAAAGCGGGAAGACTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
QY 1201 GAGCAAAACGCTGAGCCCAACATGAGAGAGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
DB 1201 GAGCAAAACGCTGAGCCCAACATGAGAGAGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
QY 1261 GGCACAGCTGACAGAGATGAGGACATGAGAGAGAGTCCGAGAGAGAGAGAGAGAGAGAG 1320
DB 1261 GGCACAGCTGACAGAGATGAGGACATGAGAGAGAGAGTCCGAGAGAGAGAGAGAGAGAGAG 1320
QY 1321 AAGGTAAGATTTGAGAGAGCAAAACCCAGCGGCAAGAAATCTCAAGATATTCAG 1380
DB 1321 AAGGTAAGATTTGAGAGAGCAAAACCCAGCGGCAAGAAATCTCAAGATATTCAG 1380
QY 1381 CCGTGTGAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
DB 1381 CCGTGTGAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
QY 1441 CAGCTGAGCTCTGCTGATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
DB 1441 CAGCTGAGCTCTGCTGATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
QY 1501 AGATTTCTAAGTTCAAGTTAAAGAAACAAACCAAGGAAAGGTTCAAGAGAG 1560
DB 1501 AGATTTCTAAGTTCAAGTTAAAGAAACAAACCAAGGAAAGGTTCAAGAGAG 1560
QY 1561 CAGAGAAAGTTCAAGTTCTCAAAATGCAATGCTTCAAGTGGAGATTTAAATCTCTGCTG 1620
DB 1561 CAGAGAAAGTTCAAGTTCTCAAAATGCAATGCTTCAAGTGGAGATTTAAATCTCTGCTG 1620
QY 1621 CAGAGAAAGTTCAAGTTCTCAAAATGCAATGCTTCAAGTGGAGATTTAAATCTCTGCTG 1680

DB 1621 CAGAGAAAGTTCAAGTTCTCAAAATGCAATGCTTCAAGTGGAGATTTAAATCTCTGCTG 1680
QY 1681 AGAGTGAAGCTTCTGAGAGAGAGGACCTCTGAGAGACAGACATCTCGGGCAAGT 1740
DB 1681 AGAGTGAAGCTTCTGAGAGAGAGGACCTCTGAGAGACAGACATCTCGGGCAAGT 1740
QY 1741 GAGCAACACTCAATGCTGTAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
DB 1741 GAGCAACACTCAATGCTGTAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1801 AACCATTGAGTAAATGATACATCAACCAAGGCTGCTTCCAGGCTCTCCATCAT 1860
DB 1801 AACCATTGAGTAAATGATACATCAACCAAGGCTGCTTCCAGGCTCTCCATCAT 1860
QY 1861 CTGGGTGCTGCTGAG 1920
DB 1861 CTGGGTGCTGCTGAG 1920
QY 1921 GGCAGACACTCACTGACAGAGAGAGAGAGATACCAAGATGCTCTGAGACCCAGGTTTC 1980
DB 1921 GGCAGACACTCACTGACAGAGAGAGAGAGATACCAAGATGCTCTGAGACCCAGGTTTC 1980
QY 1981 CTGCTGAGCTGAGAGCTCTCTGAGGTTCTTAAAGAGCTGTTAGGCTCTCATGTTG 2040
DB 1981 CTGCTGAGCTGAGAGCTCTCTGAGGTTCTTAAAGAGCTGTTAGGCTCTCATGTTG 2040
QY 2041 GAGGCAATTTCTTATTTAGGCTTGAAGGCTTGAAGATTTGAAGCTGCTCAACAGA 2100
DB 2041 GAGGCAATTTCTTATTTAGGCTTGAAGGCTTGAAGATTTGAAGCTGCTCAACAGA 2100
QY 2101 TGGACAAACATGAG 2160
DB 2101 TGGACAAACATGAG 2160
QY 2161 CAGACAAACATGAG 2220
DB 2161 CAGACAAACATGAG 2220
QY 2221 CTTCTCTTAAAGAGATTTTCTGAGTGAAGGCAATTTGAAAGAGAGAGAGAGAGAGAG 2280
DB 2221 CTTCTCTTAAAGAGATTTTCTGAGTGAAGGCAATTTGAAAGAGAGAGAGAGAGAGAG 2280
QY 2281 CAATTTGCTTAAATTTCACTACACACTACATTAACAGGTTGAATTTGAAGGTC 2340
DB 2281 CAATTTGCTTAAATTTCACTACACACTACATTAACAGGTTGAATTTGAAGGTC 2340
QY 2341 ATCTGCTCTGAGATCCCATGAGTGAATGCTGCTGACATACACTACAGAGATATTC 2400
DB 2341 ATCTGCTCTGAGATCCCATGAGTGAATGCTGCTGACATACACTACAGAGATATTC 2400
QY 2401 TATCAAAAGTGAAGATTTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2460
DB 2401 TATCAAAAGTGAAGATTTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2460
QY 2461 TGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2520
DB 2461 TGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2520
QY 2521 CATCTGTTTCACTGCTGTAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2580
DB 2521 CATCTGTTTCACTGCTGTAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2580
QY 2581 GAGATTTTCACTGCTGTAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2640
DB 2581 GAGATTTTCACTGCTGTAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2640
QY 2641 CAGAGTGAAGGAGTCTCAATGAGCTGAGCTTAAATTTCTGTAAGTGGAGGTTCC 2700
DB 2641 CAGAGTGAAGGAGTCTCAATGAGCTGAGCTTAAATTTCTGTAAGTGGAGGTTCC 2700
QY 2701 TCTGCTCCCACTGTAAGAGTGTGAATCTTACAGGAGACACTGATGATGCTGCTG 2760

Oy	2761	GCGGTGTAAGTCAAGATGTTGGACACAGCTGTACACAGTGAGCTGCAGTTTAGCCATG	2820
Db	2761	GCGGTGTAAGTCAAGATGTTGGACACAGCTGTACACAGTGAGCTGCAGTTTAGCCATG	2820
Oy	2821	GTAACTGAAGACTGCTGATCTTCTTCCTTTGTTCTTTGCTGGC	2867
Db	2821	GTAACTGAAGACTGCTGATCTTCTTCCTTTGTTCTTTGCTGGC	2867
RESULT 3			
LOCUS	MWU238332	2602 bp	mRNA linear ROD 09-Jul-1999
DEFINITION	Mus musculus	mRNA for death inducer-obliterator-1 (Dio-1).	
ACCESSION	AJ238332		
VERSION	AJ238332.1	GI:5457402	
KEYWORDS	death inducer-obliterator-1; Dio-1 gene.		
SOURCE	Mus musculus	(house mouse)	
ORGANISM	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE			
AUTHORS	1 Garcia-Domingo,D., Leonardo,E., Grandien,A., Martinez,P., Albar,J.P., Izpisua-Belmonte,J.C. and Martinez-A,C.		
TITLE	DIO-1 is a gene involved in onset of apoptosis in vitro, whose misexpression disrupts limb development		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (14), 7992-7997 (1999)		
PUBLISHED	99324176		
REFERENCE	10393935		
AUTHORS	2 (bases 1 to 2602) Garcia-Domingo,D.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-APR-1999) Garcia-Domingo D., Department of Immunology and Oncology, Centro Nacional de Biotecnologia/CSCIC, Campus Cantoblanco UAM, Ctra. de Colimemar Viejo Km 16, Madrid, 28049,		
	SPAIN		
FEATURES			
source	location/Qualifiers		
	1..2602		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10090"		
	/cell_line="MOL-1"		
	/cell_type="pre-BI"		
gene	1..2602		
	/gene="Dio-1"		
	232..>2073		
	/gene="Dio-1"		
	/function="apoptosis inducer"		
CDS	/codon_start=1		
	/evidence=experimental		
	/product="death inducer-obliterator-1"		
	/protein_id="CAA84401.1"		
	/db_xref="GI:5457403"		
	/db_xref="GOA:Q8C9B9"		
	/db_xref="GOA:Q9WV00"		
	/db_xref="SMIS:-PROT:Q8C9B9"		
	/translation="MDKGLSNREAPKAIKPTSKERKTWGPRRTTIARBGSGTS		
	ADPSBOOPQOHNLSIRSGSQPKTRVEVFITTVRRGKNPVSLEDSBPSTSV		
	TDVETLRSEGVSSSRISGPSVDLSLKEHPASSEKKVGDEEDSDSDGLTLKE		
	LQNRLRKREQEPVERSLRGSONRLRKRRREBSAFTGSVIGSAEQDRPLCKQEPNA		
	SGQPVSOSFETDIENOCVAGTNGTEENPREAKGPPECEYDPNALYCICQPHNPA		
	RPMICDRCEMFHGDQCEVIGSEARGLLBNRGDYICPNCLIQVDETNGSATNEOD		
	SGRSYVADGDTCTSIGTVKSGKGDDGIGRIEKKANPGKKKKIFQVVVARCAP		
	KCIQPGCSVIAPQPDYSYCNMDCILKHAAIMNRPLSSGEOKTKPEREVTKPRKPSIP		
	KCVGVGKIKSIVHKRLASKRENPPVKVMALARSSTSGKBACESTTPSWADHNNT		
	AVNPBRKPRTKLSPTLISKCTYHPKRGFGPGSHLDGCLGRTRLVLGVLYIVASS		
	SLPARSYQDASGPQVFLPSILMSLSGVTLSGVLMLEAISYFSFRPW"		
ORIGIN			
Query Match	90.8%; Score 2602;	DB 10;	Length 2602;
Best Local Similarity	100.0%;	Fed. No. 0;	
Matches 2602;	Conservativity 0;	Mismatches 0;	Indels 0;
		Gaps 0;	

QY	1	TCCTGTAGCTCTGGAATAATGCTGCGGATCCCGGCGCCCGGGAGCTTGTTAAGG	60
Db	1	TCCTGTAGCTCTGGAATAATGCTGCGGATCCCGGCGCCCGGGAGCTTGTTAAGG	60
QY	61	CAGTCCCACTGCCCTTGTGGTATTGGAAGCTCCGGAATCTTCTCATTTGAACTGTG	120
Db	61	CAGTCCCACTGCCCTTGTGGTATTGGAAGCTCCGGAATCTTCTCATTTGAACTGTG	120
QY	121	ACTGACACAGTGGGGGTGAGGCTTGGCCGTCTGCTTGACTGGCCCAAGTCTAATTT	180
Db	121	ACTGACACAGTGGGGGTGAGGCTTGGCCGTCTGCTTGACTGGCCCAAGTCTAATTT	180
QY	181	TATGTAGGATTTCCAGCCAAAGTTTCCAGCTTCAGTTTGAGTTTGGGACAGTATGAT	240
Db	181	TATGTAGGATTTCCAGCCAAAGTTTCCAGCTTCAGTTTGAGTTTGGGACAGTATGAT	240
QY	241	AAAGGGCACTGACCAATGAGAGACACCAAGGCTATCAAAACCACAGTAAGAGTTT	300
Db	241	AAAGGGCACTGACCAATGAGAGACACCAAGGCTATCAAAACCACAGTAAGAGTTT	300
QY	301	AGGAAACCTGGGGTTTTCCAGAGAACCAACCATTTGCCAAACGTGAGGGTGCAGAGACG	360
Db	301	AGGAAACCTGGGGTTTTCCAGAGAACCAACCATTTGCCAAACGTGAGGGTGCAGAGACG	360
QY	361	GAGGCGGACCCGAGTGGACAGCAACCAAGCAGATAACTCTCCCTGCGCGCAGTGG	420
Db	361	GAGGCGGACCCGAGTGGACAGCAACCAAGCAGATAACTCTCCCTGCGCGCAGTGG	420
QY	421	CGGACACCAAAAGCTACTGAGAGGGTGAAGAGTTTCTTACACAGTTCCGCGCCGAGGG	480
Db	421	CGGACACCAAAAGCTACTGAGAGGGTGAAGAGTTTCTTACACAGTTCCGCGCCGAGGG	480
QY	481	AAAAAGATGTGCGGGTCTCCCTGGAAGATTCCAGTGAAGCCACATCTTCCACGTCACT	540
Db	481	AAAAAGATGTGCGGGTCTCCCTGGAAGATTCCAGTGAAGCCACATCTTCCACGTCACT	540
QY	541	GATGTGAGACAGCTCCGAGGGAGCGCTTGAAGCAAGTTCTGAGATCAGAAGTGGCCCT	600
Db	541	GATGTGAGACAGCTCCGAGGGAGCGCTTGAAGCAAGTTCTGAGATCAGAAGTGGCCCT	600
QY	601	GTATCTGACTCTTAAAGGAAAGAACATCTGCGCTCTTCTGAAGGCAAAAGAGGTGA	660
Db	601	GTATCTGACTCTTAAAGGAAAGAACATCTGCGCTCTTCTGAAGGCAAAAGAGGTGA	660
QY	661	GAGGAAGAAACACTCTGACAGTGAAGTATGGCTTACGTTGAAGAACTTCAGAAC	720
Db	661	GAGGAAGAAACACTCTGACAGTGAAGTATGGCTTACGTTGAAGAACTTCAGAAC	720
QY	721	CGCTTGGAGAAAGCAGAGCAAGAACCTGTGAGAGGTTCTCTGAGAGCACTCAGAT	780
Db	721	CGCTTGGAGAAAGCAGAGCAAGAACCTGTGAGAGGTTCTCTGAGAGCACTCAGAT	780
QY	781	CGCGTGAAGAAAGCGCAGAGGAAATTTCTGCGGAACTGGGAGTGTCAATATAGC	840
Db	781	CGCGTGAAGAAAGCGCAGAGGAAATTTCTGCGGAACTGGGAGTGTCAATATAGC	840
QY	841	AGTCCCGACGAGCAAGACCTCTCTGTAAAGAGAGCTGAGGCTAGTCAAGGACCACTG	900
Db	841	AGTCCCGACGAGCAAGACCTCTCTGTAAAGAGAGCTGAGGCTAGTCAAGGACCACTG	900
QY	901	TCCCATGTACAGACAGATGACATGAAATATCATTTGGAAGGGAAGGCACTCAAGGAAAT	960
Db	901	TCCCATGTACAGACAGATGACATGAAATATCATTTGGAAGGGAAGGCACTCAAGGAAAT	960
QY	961	ACAGAGGAAACCCCAAGGAAAGCGGCAAAACAAGCTGAGTGAAGTTTACACCCC	1020
Db	961	ACAGAGGAAACCCCAAGGAAAGCGGCAAAACAAGCTGAGTGAAGTTTACACCCC	1020
QY	1021	AATGCTCTGTATCTGCAATCTGCGCCAGCTTCAACAACAGTTTATGATCTGCTGTAT	1080
Db	1021	AATGCTCTGTATCTGCAATCTGCGCCAGCTTCAACAACAGTTTATGATCTGCTGTAT	1080

QY 1081 CCGTGTGAGAGAGTGTTCATATGTCATCTGTGTGGTATTTCTGAGGCCGAGGCGGCTC 1140
 Db 1081 CCGTGTGAGAGAGTGTTCATATGTCATCTGTGTGGTATTTCTGAGGCCGAGGCGGCTC 1140
 QY 1141 CTGGAAGAGAGAGGAGAGAGTATGATCTGCGCAATTTGACCAATTTTGCAGATGAGAT 1200
 Db 1141 CTGGAAGAGAGAGGAGAGAGTATGATCTGCGCAATTTGACCAATTTTGCAGATGAGAT 1200
 QY 1201 GAGCAAAACGGTAGCGGCCCAATGAGCAGAGACTCTGTGGTGCAGATCTGTGGTGTCTGAT 1260
 Db 1201 GAGCAAAACGGTAGCGGCCCAATGAGCAGAGACTCTGTGGTGTCTGATCTGTGGTGTCTGAT 1260
 QY 1261 GGCACAGATGTCACAAAGATAGGAGAGAGTCCGAGAGAGAGAGAGAGAGAGAGAGAT 1320
 Db 1261 GGCACAGATGTCACAAAGATAGGAGAGAGTCCGAGAGAGAGAGAGAGAGAGAGAGAT 1320
 QY 1321 AAGGCTAGATTTGAT 1380
 Db 1321 AAGGCTAGATTTGAT 1380
 QY 1381 CCTGTGTGAGAGGCTCTGTGTCTCTAATGATGATGATGATGATGATGATGATGATGAT 1440
 Db 1381 CCTGTGTGAGAGGCTCTGTGTCTCTAATGATGATGATGATGATGATGATGATGATGAT 1440
 QY 1441 CAGCTTGAATCTGTGTATTTGAGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 Db 1441 CAGCTTGAATCTGTGTATTTGAGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 QY 1501 AGATTTCTAAGTTCAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
 Db 1501 AGATTTCTAAGTTCAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
 QY 1561 CAGAGAGAGTTCAGTCTTCCAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
 Db 1561 CAGAGAGAGTTCAGTCTTCCAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
 QY 1621 CACAAAGAGTTCAGTCTTCCAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
 Db 1621 CACAAAGAGTTCAGTCTTCCAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
 QY 1681 AGAGTGTGAGATCTTCTGAT 1740
 Db 1681 AGAGTGTGAGATCTTCTGAT 1740
 QY 1741 GACCAAACTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800
 Db 1741 GACCAAACTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800
 QY 1801 ACCCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860
 Db 1801 ACCCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860
 QY 1861 CTGGGTGAGTGTGAT 1920
 Db 1861 CTGGGTGAGTGTGAT 1920
 QY 1921 GGCACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
 Db 1921 GGCACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
 QY 1981 CTGGCTAAGTGTGAT 2040
 Db 1981 CTGGCTAAGTGTGAT 2040
 QY 2041 GAGGCAATTTCTAATTTGAT 2100
 Db 2041 GAGGCAATTTCTAATTTGAT 2100
 QY 2101 TGGGCAAACTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
 Db 2101 TGGGCAAACTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
 QY 2161 CAGAGCAAACTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220

Db 2161 CCAGCAGAAACATGAGAGTTTCCAGTTACTAATTTCTTCAACATGATCAGACTG 2220
 QY 2221 CTCTCTATAGAAAAGATTTTTCATGAGATGAGGCAATTTTGAAGAAGAGAGATGAG 2280
 Db 2221 CTCTCTATAGAAAAGATTTTTCATGAGATGAGGCAATTTTGAAGAAGAGAGATGAG 2280
 QY 2281 CAATTTGTCTATTTTCACTATGACACCTACATTAACAGATTTGAATTTGAAGGTC 2340
 Db 2281 CAATTTGTCTATTTTCACTATGACACCTACATTAACAGATTTGAATTTGAAGGTC 2340
 QY 2341 ATCTGTGCTCAGATCCAGATGAGTAAATGATGATGATGATGATGATGATGATGATGATGAT 2400
 Db 2341 ATCTGTGCTCAGATCCAGATGAGTAAATGATGATGATGATGATGATGATGATGATGATGAT 2400
 QY 2401 TCATCAAAAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460
 Db 2401 TCATCAAAAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460
 QY 2461 TCAGTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2520
 Db 2461 TCAGTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2520
 QY 2521 CATCTTGTTCATCTGCTGATCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2580
 Db 2521 CATCTTGTTCATCTGCTGATCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2580
 QY 2581 GTGAATTTTCAATCTGTGTTCTC 2602
 Db 2581 GTGAATTTTCAATCTGTGTTCTC 2602

RESULT 4

AK129117
 LOCUS AK129117 4906 bp mRNA linear ROD 21-NOV-2003
 DEFINITION Mus musculus mRNA for mKIAA0333 protein.

ACCESSION AK129117
 VERSION AK129117.1 GI:37359897
 KEYWORDS FLI_CDNA.
 SOURCE Mus musculus (house mouse)

ORGANISM

REFERENCE

AUTHORS

TITLE

1 Okazaki, N., Kikuno, R., Ohara, R., Inamoto, S., Koseki, H., Hirakawa, S.,
 Suga, Y., Nagase, T., Ohara, O. and Koga, H.
 Prediction of the coding sequences of mouse homologues of KIAA
 gene: III. the complete nucleotide sequences of 500 mouse
 KIAA-homologous cDNAs identified by screening of terminal sequences
 of cDNA clones randomly sampled from size-fractionated libraries

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (23-JUL-2003) Hirosaki Koga, Kazusa DNA Research
 Institute, Laboratory for Genome Informatics; 2-6-7
 Kazusa-kamata, Kisarazu, Chiba 292-0818, Japan
 (E-mail: mousekazusa.or.jp, Tel: 81-438-52-3919, Fax: 81-438-52-3918)
 The GREAT program supported by Japan science and technology
 corporation; cDNA full insert sequencing; Kazusa DNA Research
 Institute; cDNA library construction, clone selection and 5'- &
 3'-end one pass sequencing.

FEATURES

source

1..4906
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="mbg16015"
 /tissue_type="brain"
 /dev_stage="adult"
 /note="vector:modified pBC SK+"

gene
CDS

```

1. .4906
/gene="mKIA0333"
<224. .3829
/gene="mKIA0333"
/note="CDS is predicted by in silico analysis. Start codon
is not identified."
/codon_start=1
/evidence=not experimental
/product="mKIA0333 protein"
/protein_id="BAC97927.1"
/db_xref="GI:37359898"
/translation="FVGPAPGPAQFSTGTGMDKGHSNEAPALRPTSEKPKRT
MGRRTTLAKRGAGDTEVDPSHOQPOQNLILRSGVSPKRTREVEPLTVRRKG
KNVPLSDSSSTSTVDVETASGSVSSISGSRVSLGKHESSEKAGX
BEEDTSDSDGLTLKELONLRREDOPEVRLRKGREDSLETSTV
OIGSABODPLCKOPEBASOGPVOSGETDINOLEGKATQNGTENPEADGKEC
EYTDNALYCTCRQPHNRFNCCDRCEPHDCTGISIBARLLERGEDYICNC
TLYQVDETNSATBDSDGCRSVGADGTCSTIGVEQSGEDQIKGIRKANPS
GRRUKITPQVVPAPGAPKCIQPCSSVAPDSVCSNDCILKHAATVPLSGKEQ
KTRPKKVTKEKESLPCKSVOVGKISVHRLASRGRNPKVLMLASSETGK
EACSSSTPSMA5DHYNVAVKEPKPTPLSPITLSKMDRVRVEDTMAVTLPK
KALPSASLVGRQTSPPNIVPKLIPYSNMAAKPAIKKLPQSGKTIIPRPMPSATLS
GTSARQAGTPMTTAAKSLPGSAAVYGTTRKMSANVPAAPRGLGVPSPAPSPBN
SQRNTRSLKSLILMKRVNDSDDLMTNENVKLALIEKEMFNLPOYTDNRKSKY
KSLMFNLKDPKNOGLFHRVLAREBISLAKVNRKPEELVERKLSMTKEPTKSIISRT
RLINENKNTTPEETIPDEMDSPVSDSEQBSVRAAPERKSNAPLLDVFSMLDOTT
SGHRAHLPDANKICTGVPSSESDPAKQKLSASKEDFPRSDSPNAPVPTA
DEGIADTPEKNSSEDPBSTSLNORKEFPSPSDSHPEPSLGLSPSSASGSGV
VTTVMGSDPRTALSGCTVYASMAAHJNDSQASPTKDMKRALTSVAPPSILAK
PSSSDPRLTSPSPSPSIESRSPPREGDTTLPLSLRNTIKGPIINQSAKPYTKAY
VSGCDYLSIEDLPDTHIGRLAPKTVMVYGLKSSVSKELLLRPHATEBEREAY
ISLVSYSFSGRGRGVANNRHHVXD.YLLPLSAKDPVPSKLPFBGPGKPVSGR"

```

ORIGIN

Query Match 61.8%; Score 1770.4; DB 10; Length 4906;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1774; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

44 GGAGCTGTTTAAAGAGCAGTCCCACTGTCCTTGTGTGTGCAAGCTCCGAATCTT 103
90 GAGCTGTTTAAAGAGCAGTCCCACTGTCCTTGTGTGTGCAAGCTCCGAATCTT 149
104 CTCAATTGATGAACCTGTGACCAACAGTGGGGTGAAGCTTGGCCCTTCTGCTTGA 163
150 CTCAATTGATGAACCTGTGACCAACAGTGGGGTGAAGCTTGGCCCTTCTGCTTGA 209
164 CCCCAAGCTCTAATATTTTATGTAGATTTCCAGCCCAAGTTTCCAGCTTTCAGTTTG 223
210 CCCCAAGCTCTAATATTTTATGTAGATTTCCAGCCCAAGTTTCCAGCTTTCAGTTTG 269
224 GGAAGATGTGATGATTAAGGCACTTGAAGCAATGAGAAAGCAAGCTATCAAC 283
270 GGAAGATGTGATGATTAAGGCACTTGAAGCAATGAGAAAGCAAGCTATCAAC 329
284 CCACCAAGTAAAGAGTTCAAGAAAACTTGAGGTTTGAAGAACCAAGATTCGAAACGTG 343
330 CCACCAAGTAAAGAGTTCAAGAAAACTTGAGGTTTGAAGAACCAAGATTCGAAACGTG 389
344 AGGGTCAGGAACACAGAGGGGGAACCCCAAGTGAAGAGAACCAAGAGCATTAACCTCT 403
390 AGGGTCAGGAACACAGAGGGGGAACCCCAAGTGAAGAGAACCAAGAGCATTAACCTCT 449
404 CCCTGCGCGCAGTGAACGCAACCAAAAGTAACTGAGAGGGTGAAGAGTTCTTAACA 463
450 CCCTGCGCGCAGTGAACGCAACCAAAAGTAACTGAGAGGGTGAAGAGTTCTTAACA 509
464 CGGTTTCGCGCGCAGAGGAAAAAAGATGTCCCGGTGTCCCTGAGAGATTCAGTGAGCCCA 523
510 CGGTTTCGCGCGCAGAGGAAAAAAGATGTCCCGGTGTCCCTGAGAGATTCAGTGAGCCCA 569
524 CATCTTCCACATCTGATGTGTGAGACAGCTTCCGAGGGGAGAGCTGTGAAGAGATTTCTG 583
570 CATCTTCCACATCTGATGTGTGAGACAGCTTCCGAGGGGAGAGCTGTGAAGAGATTTCTG 629

```

```

584 AGATCAGAAAGTGGCCCTGTATCTGACTCTTAAAGGAAAGAACTTCCTGCTTTCTGAAA 643
630 AGATCAGAAAGTGGCCCTGTATCTGACTCTTAAAGGAAAGAACTTCCTGCTTTCTGAAA 689
644 AGGCAAAAGAGGTGAAGAGGAAAGAAAGCACTCTGACAGTGAACGTAAGGCTTAAGT 703
690 AGGCAAAAGAGGTGAAGAGGAAAGAAAGCACTCTGACAGTGAACGTAAGGCTTAAGT 749
704 TGAAGAACTTCAAGAACCGCTTTCGAGAAAGCGGAGCAAGAACTGTGAGAGTCCC 763
750 TGAAGAACTTCAAGAACCGCTTTCGAGAAAGCGGAGCAAGAACTGTGAGAGTCCC 809
764 TGAAGAGCAGTCAAGATTCGCTGAGAAAGAGCGGAGCAAGAACTTCGCGAAAATG 823
810 TGAAGAGCAGTCAAGATTCGCTGAGAAAGAGCGGAGCAAGAACTTCGCGAAAATG 869
824 GGAAGTCCAAATAGGCAAGTCCGAGACGAGACCTCTCTGTTAAGCAGAGCTTGAAG 883
870 GGAAGTCCAAATAGGCAAGTCCGAGACGAGACCTCTCTGTTAAGCAGAGCTTGAAG 929
884 CTAGTCAGGGAACAGTATCCAGTCAAGACAGATGACATAGAAATCACTTGAAGAGGA 943
930 CTAGTCAGGGAACAGTATCCAGTCAAGACAGATGACATAGAAATCACTTGAAGAGGA 989
944 AGGCGAATCAAGGAAATACAGAGGAAATCCCAAGGAAAGCGGCAAACTCAAGT 1003
990 AGGCGAATCAAGGAAATACAGAGGAAATCCCAAGGAAAGCGGCAAACTCAAGT 1049
1004 GTGAGTTTAAAGCAACCCCAATTCGCTTACATGCAATCTGCGGCAAGCTTCAAGCA 1063
1050 GTGAGTTTAAAGCAACCCCAATTCGCTTACATGCAATCTGCGGCAAGCTTCAAGCA 1109
1064 TTATGATCTGCTGTGATTCGAGTGAAGAGTGTTCATAGTGTGCTGTGCTTCTG 1123
1110 TTATGATCTGCTGTGATTCGAGTGAAGAGTGTTCATAGTGTGCTGTGCTTCTG 1169
1124 AGGCGCAAGGCGGCTCTTGAAGAAAGCAAGGGAAGATCAATCTGCGCAAAATGCA 1183
1170 AGGCGCAAGGCGGCTCTTGAAGAAAGCAAGGGAAGATCAATCTGCGCAAAATGCA 1229
1184 TTTTGAAGTGAAGATTAAGACAAACGCTAGCCCAATGAGCAGAACTCTGCGTGA 1243
1230 TTTTGAAGTGAAGATTAAGACAAACGCTAGCCCAATGAGCAGAACTCTGCGTGA 1289
1244 GATCTGTGGGTGTGATGTGACAGATGCAAGCAATGAGGACAGTGAAGCAGAGTCCG 1303
1290 GATCTGTGGGTGTGATGTGACAGATGCAAGCAATGAGGACAGTGAAGCAGAGTCCG 1349
1304 GAGAAACCAAGGCAATTAAGGATTAAGAAAGCAAGCAACCCCAAGGCAAGAAAA 1363
1350 GAGAAACCAAGGCAATTAAGGATTAAGAAAGCAAGCAACCCCAAGGCAAGAAAA 1409
1364 AACTCAAGATTAATTCAGGCTGTGTGTAAGAGCTCTTGTGTCTTAATGATAGTCCG 1423
1410 AACTCAAGATTAATTCAGGCTGTGTGTAAGAGCTCTTGTGTCTTAATGATAGTCCG 1469
1424 GGTGTTCAGTGTGAACAGGCTTGACTGTGTATGAGATTAATGATAGTGTCTCAAC 1483
1470 GGTGTTCAGTGTGAACAGGCTTGACTGTGTATGAGATTAATGATAGTGTCTCAAC 1529
1484 ACGCAGACGTAACATGAGATTTCTAAGTTCAAGTAAAGAAACCAAAACCAACG 1543
1530 ACGCAGACGTAACATGAGATTTCTAAGTTCAAGTAAAGAAACCAAAACCAACG 1589
1544 AAAAGTCAAGACGAAGCGGAAAGAAAGTTCAAGTTTCAAAATGCAAGTTCAGTGGGA 1603
1590 AAAAGTCAAGACGAAGCGGAAAGAAAGTTCAAGTTTCAAAATGCAAGTTCAGTGGGA 1649
1604 TTAATATCTCTTCTGTGCAAGAGACTAGCGTCAAGAAAGAGGAAACCCAGTGAAGA 1663
1650 TTAATATCTCTTCTGTGCAAGAGACTAGCGTCAAGAAAGAGGAAACCCAGTGAAGA 1709

```

0y	1664	AAGGAGAGCTGAGCTTCCAGAGAGTGAAGCTTCTGGGAAAGAGGACGCTGGAGAGCAGCA	1723
Db	1710	AAGGAGAGCTGAGCTTCCAGAGAGTGAAGCTTCTGGGAAAGAGGACGCTTGGAGAGCAGCA	1769
0y	1724	CACCATCTCGGAGCAAGTGAACCAACAACACTACATATGCTGTGAAGCCAGAGAGCCAGAGAAGC	1783
Db	1770	CACCATCTCGGAGCAAGTGAACCAACAACACTACATATGCTGTGAAGCCAGAGAAGC	1829
0y	1784	CCACTGACATCTGCGCCACCCCTATTATGATTAATGTACGTA	1823
Db	1830	CCACTGACATCTGCGCCACCCCTATTATGATTAATGTACGTA	1869
RESULT 5			
AL732560/c			
LOCUS	AL732560	104663 bp	DNA linear ROD 22-AUG-2002
DEFINITION	Mouse DNA sequence from clone RP24-209P15 on chromosome 2, complete		
ACCESSION	AL732560		
VERSION	AL732560.11		
KEYWORDS	HTG.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 104663)		
TITLE	Leonamornlert,D.		
JOURNAL	Direct Submission		
COMMENT	Submitted (22-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Aug 23, 2002 this sequence version replaced gi:22204532.		
	----- Genome Center		
	Center: Wellcome Trust Sanger Institute		
	Center code: SC		
	Web site: http://www.sanger.ac.uk		
	Contact: humquery@sanger.ac.uk		

	During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.		
	This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one MJ3 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWSNPSPOT; Tr: TRMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at		
	http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP24-209P15 is from a Male (C57BL/6J) mouse BAC Library VECTOR: pTARBAC1.		
FEATURES			
source	Location/Qualifiers		
	1..104663		
	/organism="Mus musculus"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:10090"		
	/chromosome="2"		
	/clone="RP24-209P15"		
	/clone_1fb="RPCT-24"		
ORIGIN			
Query Match	43.3%	Score 1241.2;	DB 10; Length 104663;
Best local Similarity	99.1%;	Pred. No. 0;	
Matches 1269;	Conservative	0; Mismatches	8; Indels 3; Gaps 2;
1588	AGGTTCAGTGGAGATTAAATCTCTTCTGTGCAAGAGACTAGCTCAGAGAAAAG	1647	

Db	46536	ATTATCCAGGTGGGAGTTTAAATCTCTTCGTGCACAAGACTGACGTCACAGAAAAGG	462717
OY	1648	GAAAACCTCAGTGAAGAAAGTGAATGCTGCGCTTCCAGAGTGAACATTTCTGGGAGAGAGCA	1707
Db	46276	GAAAACCTCAGTGAAGAAAGTGAATGCTGCGCTTCCAGAGTGAACATTTCTGGGAGAGAGCA	46217
OY	1708	GCCCTGAGAGCAGACACACATCTCTGGGCAAGTACACAACACTCAATCTCTGAAGCCA	1767
Db	46216	GCCCTGAGAGCAGACACACATCTCTGGGCAAGTACACAACACTCAATCTCTGAAGCCA	46157
OY	1768	GAGAAACCAGAGAGGCCCACTGCACTCTGCGCCACCTATTGAGTAAATGTAAGTATCAC	1827
Db	46156	GAGAAACCAGAGAGGCCCACTGCACTCTGCGCCACCTATTGAGTAAATGTAAGTATCAC	46097
OY	1828	CCAAAAGCTGGCTTCCAGGAGCCCTCCCATCAATCTGGGAGGCGCTGGGGGCTGTCTAGG	1887
Db	46096	CCAAAAGCTGGCTTCCAGGAGCCCTCCCATCAATCTGGGAGGCGCTGGGGGCTGTCTAGG	46037
OY	1888	ACCAGAGTCTCTGGGAGTTCCTGAGTCAATAGTACAGAGCACTCACTGCACCCAGAAC	1947
Db	46036	ACCAGAGTCTCTGGGAGTTCCTGAGTCAATAGTACAGAGCACTCACTGCACCCAGAAC	45977
OY	1948	AGATACCAAGATGCTCTTGAACCCCAAGTGTTCCTGGCCTAAGCTGTGAGAGCCCTCTGGG	2007
Db	45976	AGATACCAAGATGCTCTTGAACCCCAAGTGTTCCTGGCCTAAGCTGTGAGAGCCCTCTGGG	45917
OY	2008	TGTTTCTTAAGAGCTGTGAGAGCCTCATAGTTGGAAGGCAATTTCTATTTCAGTTTTAGG	2067
Db	45916	TGTTTCTTAAGAGCTGTGAGAGCCTCATAGTTGGAAGGCAATTTCTATTTCAGTTTTAGG	4585
OY	2068	CTTGTGTAAGTATTGAACTGCTCTCAACAAGTGGGACAACATCAGTGCACATCTTCAAA	2127
Db	45856	CTTGTGTAAGTATTGAACTGCTCTCAACAAGTGGGACAACATCAGTGCACATCTTCAAA	45797
OY	2128	CTCAGTTCACACTTGAAGTTCCTCAACAGACAGCCAGCACAAAACATGAGTTTTCCAC	2187
Db	45796	CTCAGTTCACACTTGAAGTTCCTCAACAGACAGCCAGCACAAAACATGAGTTTTCCAC	45736
OY	2188	GTTACCTATTATTTCTTCAACATGAGTCAACACTGCTCTTCTATTAGAAAAGTATTTTTTC	2247
Db	45737	GTTACCTATTATTTCTTCAACATGAGTCAACACTGCTCTTCTATTAGAAAAGTATTTTTTC	45686
OY	2248	ATGAGTTAAGGCAATTTTGAAGAAGACAGTGAACAATTTGTGTCTATTTCAACATCTACA	2307
Db	45679	ATGAGTTAAGGCAATTTTGAAGAAGACAGTGAACAATTTGTGTCTATTTCAACATCTACA	45622
OY	2308	CACCTACATAAACAAGTGTGAATTTGAATTTGAAGTCAATCTGCTGCACAGTCCCATGTGAGT	2367
Db	45619	CACCTACATAAACAAGTGTGAATTTGAATTTGAAGTCAATCTGCTGCACAGTCCCATGTGAGT	45566
OY	2368	AATGTGTGTACCACTACACTTACAGACATATTTCTATCAAAAGTGAATAATGTGGGTGGA	2427
Db	45559	AATGTGTGTACCACTACACTTACAGACATATTTCTATCAAAAGTGAATAATGTGGGTGGA	45506
OY	2428	GCTTCCAGAGCTGTGAGATGCCAGATTTTCAAAATCAAGTGTATTATGTTTGAAGTAACTTT	2487
Db	45499	GCTTCCAGAGCTGTGAGATGCCAGATTTTCAAAATCAAGTGTATTATGTTTGAAGTAACTTT	45446
OY	2488	TGAAACCAGATTAGATCACTGAGTTTGAAGCAATTTCTTGTTCACCTGCTGTACAAGT	2547
Db	45439	TGAAACCAGATTAGATCACTGAGTTTGAAGCAATTTCTTGTTCACCTGCTGTACAAGT	45386
OY	2548	AACACTTCTCTATCTTTTGTACAGTGTGTTCAAGTAAATTTTCACTGTGTCTCATGAT	2607
Db	45379	AACACTTCTCTATCTTTTGTACAGTGTGTTCAAGTAAATTTTCACTGTGTCTCATGAT	45326
OY	2608	AGTGTATTTGTTCATCTGTGTCAACCTCAAGCCCAAGATTAAGGAGGTCTCATGGCCTAAG	2667
Db	45319	AGTGTATTTGTTCATCTGTGTCAACCTCAAGCCCAAGATTAAGGAGGTCTCATGGCCTAAG	45266
OY	2668	CTGGCCTTAAATTTTCTGTAAATGTGGGGGTCTCTCTGCCCCCACTGTAAAGAGTCTG	2727
Db	45259	CTGGCCTTAAATTTTCTGTAAATGTGGGGGTCTCTCTGCCCCCACTGTAAAGAGTCTG	45206

2728 AACCTTACGAGCAGACACTGATGTATGCTGTGGGTGTGAACCTAGAGTGTGGACCG 2787
|||||
45199 AAACCTTACGAGCAGACACTGATGTATGCTGTGGGTGTGAACCTAGAGTGTGGACCG 45140
|||||
2788 CAGTCTTACGAGCAGTGTGAGTGTGCTTGTAGCCATGTGTAAAGTGTGTGCTTGTG 2847
|||||
45139 CAGTCTTACGAGCAGTGTGAGTGTGCTTGTAGCCATGTGTAAAGTGTGTGCTTGTG 45080
|||||
2848 TTTGTTCTTCTTGTGTGCTGGG 2867
|||||
45079 TTTGTTCTTCTTGTGTGCTGGG 45060
|||||
RESULT 6
BD244712 2610 bp DNA linear PAT 17-JUL-2003
LOCUS Genes encoding for the human and murine death
DEFINITION inducer-oblierator-1.
ACCESSION BD244712
BD244712.1 GI:33054482
VERSION JP 2002526040-A/1.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 2610)
AUTHORS Alonso, C.M., Domingo, D.G., Grandien, A., Leonardo, B. and Martinez, P.
TITLE Genes encoding for the human and murine death inducer-oblierator-1.
JOURNAL Patent: JP 2002526040-A 1 20-AUG-2002;
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS
COMMENT OS Homo sapiens (human)
PN JP 2002526040-A/1
PD 20-AUG-2002
PF 10-SEP-1999 JP 2000570314
PR 10-SEP-1998 SE 9803069-5, 17-SEP-1998 US 60/100873 PI
CARLOS MARTINEZ ALONSO, DAVID GARCIA DOMINGO, ALF GRANDIEN, PI
ESTHER LEONARDO,
PI PEDRO MARTINEZ
PC C12N15/09, A61K38/00, A61K45/00, A61K48/00, A61P3/10, A61P17/00, PC
A61P29/00,
PC A61P35/00, A61P37/06, A61P43/00, C07K14/47, C07K16/18, C12N1/21, PC
C12N5/10,
PC C12P21/02, C12Q1/02, C12Q1/68, G01N33/15, G01N33/50//C12P21/08, PC
C12N15/00,
PC C12N5/00, A61K37/02
CC This gene is referred to in the application as Human Death
Inducer
CC Obliterator Gene 1.
CC It has now been named by the Human Gene
Nomenclature Committee
CC Death-Associated Transcription Factor (DATE-1) FH Key
Location/Qualifiers
FT source 1..2610
location/Qualifiers
1..2610
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 35.7%; Score 1023.6; DB 6; Length 2610;
Best Local Similarity 71.9%; Pred. No. 3.7e-303;
Matches 1494; Conservative 0; Mismatches 534; Indels 50; Gaps 10;

QY 1 TCCGTGTAGCTCTGTGAAATGCTGTGAGATCCCGCGCGCGGAGCTTTGTTAAGAG 60
Db 21 TCCGGGCGCTTGGGGAAATGGCTGTGAGACCTTAAAGGCTTGGAGCTTACTCACA 80
QY 61 CAGTCCCACTGTCCCTTGTGTGTGCAAGCTCCGGAATCTTCTCATTAAGAACTGTG 120
|||||

Db 81 GAACAGCCTCTAATATATCTGAATGTGTGAAAAATACAAAGCCTGTTACTGTGAACATG 140
QY 121 ACTGACAAACAGTGGGGTGAAG--GCTTGGCCGTCTGCTTGAAC-----TGGCCC 166
Db 141 GGTGACAAACAGTGTGTGTGTGTGAGCTTGGCTGTCTGTTGTGACCCAGAGGTTGTCTG 200
QY 167 CAGGTCTATTAATTTTATATGTAAGATTTCCAGCAAGGTTTCCAGCTTTCAGTTTGGGA 226
Db 201 AGGGTTTGTGTATTTAATTAAGATTT-CAGGGAAAAAGTGTCCAGCTTTCAGTGTGTGAG 259
QY 227 CAGGTATGATATGATTAAGGCACTGAGCAATGAGAAACCAAGGCTATCAAAACCA 286
Db 260 CAGGTATGACACAAAGGCAACCCGAGCAATGAGAGGACCTTAAGGCAATCAAAACCA 319
QY 287 CAGGTATGAGTTCCAGAAATCTGGGGTTTCCAGAAACCAAGTTCGCAAAAGTGAAG 346
Db 320 CAGCAAAAGATTCAGAAATCATGGGGTTTCCAGAAACCAAGTTCGCAAAAGTGAAG 379
QY 347 GTGCAAGACACAGAGGCGAACCCCATGTAGCAGCAACCC-----ACAGCAATTAAC 400
Db 380 GCGCAGGGGACCGCGAGGCTGAACCCCATGTAGCAGCAACCCCATGTAGCAGCTGGGCC 439
QY 401 TCTCCCTGCGCGCAGTGTGACGCAACCAAAAGTATCTGAGAGGTTAAGAGATTTCTTA 460
Db 440 TGTCCCTGTGCGCGCAGTGTGAGGCAACCCAGGCACTGAGGCGCTGTGAGCAAGTTCTGA 499
QY 461 CCAAGGTTGCGCGCGAGGAAAGAAATGTGCGGTTGCTTCCAGAGATTCAGTGAAG 520
Db 500 CATTGTGCGGGGCGCGCGCAGAGAGCAATGCTTCTTCCGTGAGAGATTCGTGTAGC 559
QY 521 CCAATCTTCCACAGTCACTGATGTGTGAGCAAGCTTCCAGAGGAGCGTTGAAGACATT 580
Db 560 CCAAGTCTGCGCGCGCAGAGAGCGCGAGCAAGCTTCCAGAGGAGCGTTGAAGAGCGCTT 619
QY 581 CTGAGATTCAGAGTGGCCCTGTATCTGACTCTTA---GGAAAGAAACATCTGCTCTT 637
Db 620 CTGAGACCAAGAGCGCGCCCAAGTGTGCTTCCACACTGTGAAGAGCAACAGCCCTCTT 679
QY 638 CTGAAAGGCAAAAGAGGTGAAGAGAAAGACACTCTGCAAGTGAAGTGAAGGCG 697
Db 680 CTGAAAGGTTGAAGAGGAGATGACACGATGACCTTCCATATGTGACAGAGATGACC 739
QY 698 TTACGTTGAAGAACTTCAAGAACCGCTTCCAGAAAGCGAGAGCAAGACTGTGAGAGA 757
Db 740 TGACTTTGAAGAGCTTCAAGATGCGCTTCCAGAGAGGCGAGAACAGACCCACGTAGA 799
QY 758 GTTCCCTGAGAGGCACTGAGATGCGCTTGAAGAAAGCGCAGAGAGAAAGATTCTGCC 817
Db 800 GGCCTCTGAAGGAGATCCAGAGTCCCTGCGAAGAGAGCGCCGAGAGAGGTTCCGCC 859
QY 818 AAATCGGAGTGTCCAAATAGGCAATGCGAGAGGACAGACCTCTCTTGAAGAGAGC 877
Db 860 AGACTGTGGGCTTCCAGAGCGAGTGAACATGTGTGAGAGGCGTCTGCGCAGTGAAGAGAGC 919
QY 878 CTGAGGCTAGTCAAGGACAGTGTCCAGTCAAGAGCAAGTGAATAGTGAATCAATGTTG 937
Db 920 CCGAAGCAATCAAGGCGGTTGTGTCCAGGCTGGAAGATGAACAGAGATGATGATTTGG 979
QY 938 AAGGAGAGGCACTCAAGGAAATTAAGAGAAATCCCAAGGAGCGGCAAAACCAAGC 997
Db 980 AAGGAAAGCGGCTCAAGACATCAAGATGAGAGGCTGAGAGCTTGTGGCGCAGCGAAGC 1039
QY 998 TTTGTTGAGGTTTACGACCCCAATGCGCTGTAAGTCAATGCGCGCGAGCTCAACA 1057
Db 1040 CTGAATGTAGAGTTACGACCCCAAGCGCTGTATGTGCAATTTGCGCGAGCTCAACA 1099
QY 1058 ACAGGTTATGATCTGCTGTGATCGGTGTGAGAGAGTGTCCATGATGATCTGTGTGGA 1117
Db 1100 ACAGGTTATGATTTCTGTGACCGCTGTGAAGAAATGTTTCAATGCAATGTGTGGGCA 1159
QY 1118 TTTCTGAGGCGGAGGCGGCTCTCTGAAAGAAAGCGGAGATCAATCTGCCCAATT 1177
Db 1160 TTTCTGAGGCTGAGAGGAGGCTTTTGAAGAAAGATGAGGAGATATATCTGCCCAACT 1219
|||||

```

QY 1178 GCACATTTTGCAGTGAAGATGAGAAACGGTAGGCCCAATGAGAGAGCTCTG 1237
DB 1220 GCACATTTTGCAGTGAAGATGAGAAACGGTAGGCCCAATGAGAGAGCTCTG 1279
QY 1238 GGTGAGATCTGTGGGTCTGTATGACAGACTGCAAGAGATGAGAGAGAGAGAG 1297
DB 1280 AATGAGAGCTGTGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1339
QY 1298 AGTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357
DB 1340 AGTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1399
QY 1358 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1417
DB 1400 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1459
QY 1418 GCGCTGGGTGTTCAGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1477
DB 1460 GCGCTGGGTGTTCAGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1519
QY 1478 TCAAAACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1537
DB 1520 TCAAAACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1579
QY 1538 CCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1597
DB 1580 CCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1639
QY 1598 TGGGATTAATTAATCTCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1657
DB 1640 TGGGATTAATTAATCTCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1699
QY 1658 TGAAGAA---AGTATGCTGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAG 1714
DB 1700 TGAAGAAAGCAGTGTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1759
QY 1715 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1774
DB 1760 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1819
QY 1775 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1834
DB 1820 CAG---CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1870
QY 1835 CTGGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1889
DB 1871 TTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1930
QY 1880 CAG-----AGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1940
DB 1931 TGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1990
QY 1941 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2000
DB 1991 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2050
QY 2001 CTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2038
DB 2051 TCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2088

```

```

RESULT 7
AX023369 2610 bp DNA linear PAT 15-SEP-2000
LOCUS AX023369 1 from Patent WO0015787.
DEFINITION AX023369
ACCESSION AX023369
VERSION AX023369.1 GI:10183781
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE
1
AUTHORS Leonardo, E., Martinez, P., Alonso, C.M., Domingo, D.G. and Grandien, A.
TITLE Genes encoding for the human and murine death in
ducer-obliterator-1
JOURNAL Patent: WO 0015787-A 1 23-MAR-2000;
LEONARDO ESTHER (ES) ; MARTINEZ PEDRO (ES) ; CONSEJO SUPERIOR
INVESTIGACION (ES) ; DOMINGO DAVID GARCIA (ES) ; MARTINEZ ALONSO
CARLOS (ES) ; BANERMAN DAVID GARDNER (GB) ; GRANDIEN ALF (SB)
FEATURES
source
1..2610
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="This gene is referred to in the application as
Human Death Inducer Obliterator Gene 1. It has now been
named by the Human Gene Nomenclature Committee as
Death-Associated Transcription Factor (DNF-1)"
ORIGIN
Query Match 35.7%; Score 1023.6; DB 6; Length 2610;
Best Local Similarity 71.9%; Pred. No. 3.7e-303;
Matches 1494; Conservative 0; Mismatches 534; Indels 50; Gaps 10;
QY 1 TCCGTGTAGCTCTGGAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
DB 21 TCCGTGTAGCTCTGGAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 80
QY 61 CAGTCCCACTGCTCCCTTGTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 81 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 140
QY 121 ACTGACACAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 166
DB 141 GCTGACACAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 200
QY 167 CAGTCTATATTTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 226
DB 201 AGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 259
QY 227 CAGGTATGATATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 286
DB 260 CAGGTATGATATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 319
QY 287 CAGGTATGATATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 346
DB 320 CAGGTATGATATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 379
QY 347 GTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 400
DB 380 GTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 439
QY 401 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 460
DB 440 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 499
QY 461 CCAAGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 520
DB 500 CCAAGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 559
QY 521 CCAAGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 580
DB 560 CCAAGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 619
QY 581 CTGAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 637
DB 620 CTGAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 679
QY 638 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 697
DB 680 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 739
QY 698 TTACGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 757

```


Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMW at: <http://image.llnl.gov>
Series: IRAL Plate: 2 Row: 0 Column: 16
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 18375620.

FEATURES

SOURCE

1. 2767
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:896 IMAGE:2959493"
/tissue_type="Colon, adenocarcinoma"
/clone_id="NIH MGC_15"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1. 2767
/gene="DATP1"
/note="synonyms: D101, FLJ11265, KIAA0333"
/db_xref="locusid:11083"
282. 1970
/codon_start=1
/product="death associated transcription factor 1, isoform a"
/protein_id="AAH04237.1"
/db_xref="gi:13278981"
/db_xref="locusid:11083"
translation="MDDKGDPSNEBAPAIKPTKFRKTMGFRRTTIARREGADAE
PATDAETBSBSVBSASSTRSGPQASRAVREPPASISRVKGGSDHDSDGLT
LKEIDNRIRKREDEPTERPLKIOSRLAKRREBPATVBSDYETVLPKOR
PENQGVVSQAKGKDESLBKAADI KDEPGLGRRPECEYDPAALCTICOP
HNNRPMICDRCEWPHGCVGISBARGLIERNGSDYICPNTLIQVDETSBAD
QOBAKRMFGDADGTDCTSIGTEKSGSDQIKGRIEKAANSCKKXKIPQVIBAP
GASKICPGCCCAVADPSVYSGNDCLIKHAATMKELSGKQKPEKSKMKPEK
SLPKGADAGIKTISVHRPAREKETTKYKAVVVPASSEALGKAEGSSPSMAD
HNNRPMFKTAPSPSLYKCKMTHGLGLDPSRSPWAIIPWACPGALALC"
1092. 1241
/note="PHD; Region: PHD-finger. PHD folds into an
interleaved type of Zn-finger chelating 2 Zn ions in a
similar manner to that of the RING and FYVE domains"
/db_xref="cdm:pfam00628"

CDS

gene

misc_feature

1092. 1241
/note="PHD; Region: PHD-finger. PHD folds into an
interleaved type of Zn-finger chelating 2 Zn ions in a
similar manner to that of the RING and FYVE domains"
/db_xref="cdm:pfam00628"

ORIGIN

Query Match 35.4%; Score 1016.2; DB 9; Length 2767;
Best Local Similarity 73.5%; Pred. No. 7, 2e-301;
Matches 1414; Conservative 0; Mismatches 474; Indels 36; Gaps 8;

141 GCTTGCCGCTCTGCTTGAACCTGACCTGACCTGATTAATTATTTATGATGATTTCCAGCCAA 200
192 GCTTGACCCACAGAGGTTCTGCTGCGCAGGGTCTTGTGTGTTAGATTT-CAGGGAA 250
201 AGTTTCAAGCTTCAGTTTGGGACAGGTATGATGATTAAGGGCACTGAGCAATGA 260
251 AAGTGTCCAGCTTTCAGTTGAGACAGGTATGACGACAAAGGCACTCCAGCATGA 310
261 GGAAACACCCCAAGGCTATCAACCCACAGTAAGAGTTCAAGAAAACTGGGGTTTTG 320
311 GGAGGACACTAAGGCATCAAAACCCACGAAAGAGTTCAAGAAAACTGGGGTTTTG 370
321 AAGAACCAAGATTGCAAACTGAGAGGTGACAGAGACAGGAGCGGACCCAGAGAGA 380
371 AAGGACCACTATCGCCAAAGAGAGGCGCAGAGGACGCGAGAGCTGACCCACTGAGCC 430
381 GCAACCC-----ACAGACGATTAACCTCTCCCTGCGCGCAGTGAAGGCAACCAAAAG 434
431 GCCACCCCAACAGACAGAGCTGGGCTGTCTCTGCGCGCAGTGAAGGCAACCCAAAGG 490
435 TACTGAGAGGTGAAGAGTTTCTTACCAAGGTTGGCGCCCGAAGGAAAAAGATGTGCC 494
491 CACTGAGCGCGTGAAGAGTCTCTGACCAATTCGCGCGCGCGCGAGAGAGAGATGCC 550

495 GGTGTCCTGAGAGATTCAGTAGAGCCCACTTCCACAGTCACTGATGTGAGACAGC 554
551 TGTCTCTCTGAGAGATTCGTGTGAGCCACGCTCTGCCCCGACAGACGCCGAGACAGC 610
555 TTCCGAGGGAGCGTTGAAAGAGTTCGAGATCAGAAAGTGGCCCTGTATCTGACTCTT 614
611 CTCCGAGGGCAGCGTGAAGAGCGTTTGAAGACAGAAAGCGGCCCTCACTGTCTTCCAC 670
615 A---GGGAAAGAACTCTGCTCTTTTGGAAAAGGCAAAAGGAGTGAAGAGAGAA 671
671 AGCTGTGAAGAAACACACAGCTCTTGTGAAAAGGTGAAGAGGAGATGACCAATGA 730
672 CACCTCTGACAGTGAACATGATAGCTTACGTTGAAGAACTTCAAGACCGCTTCGAG 731
731 CACCTCCATGATGACAGGATGCGCTGACCTTGAAGAGCTTCAAGAAATGCGCTTCGAG 790
732 AAAGCGAGCAAGAACTGTGAGAGGTCCCTGAAGGCACTCAGAAATGCGCTTGAAGAA 791
791 GAAGCGGAAACAGAGGCCCACTGAGAGGCCCTGAAGGGAATCCAGAGTGGCTGCGGAA 850
792 GAAGCGCAGAGAGAGAGATTTCTGCGGAAACGAGGATGTCCAAATGAGAGTCCGAGCA 851
851 GAAGCGCGGAGAGAGGTCCTCCGAGACATGTGGGCTTCCGAGCCAGTGAACATGTGGA 910
852 GGACAGACCTCTCTGTAAGCAGAGAGCTGAGGCTAGTCAAGGACCAAGTGTCCAGTCA 911
911 GGCGCTCTGCGCCAGTGAAGAGAGCCCGAGAAACGATCAAGGGGTTGTCTCCAGGCTGG 970
912 GACAGATGACATGAAATACATTGGAAGAGAGGCGCATCAAGGAAATACAGAGAAA 971
971 GAAAGATGACAGAGAGATGATGAGAGGAAAGCGGCTCAGGACATCAAGATAGAGA 1030
972 CCCGAGGAGAGCGGCAAAACCAAGCTGAGAGTGAAGTTTACACCCCAATGCCCTGTA 1031
1031 GCTGGAAGCTTGGGCGGACCGAAGCTGAAGTGAAGGTTTACGACCCCAACCCCTGTA 1090
1032 CTGCACTGCGCCGCAAGCTCAACAACAGTTTATGATCTGCTGATCGGTGAGGA 1091
1091 TTGCAATTGCCCGCAGGCTCAACAACAGTTTATGATTTGCTGTGACCGCTGTGAAG 1150
1092 GTGGTTCCATGTGACCTGTGGGATTTTCTGAGGCCGAGGCGGCTCTGGAAGGAA 1151
1151 ATGGTTTCATGGCGATGTTGTGTGGCATTTCTAGGGCTCAGAGGAGGCTTTTGAAGGAA 1210
1152 CGGGAGAGCTACATCTGCCCAATTTGCAACATTTTGCAGGTGACAGATGAGCAACCG 1211
1211 TGGGAAAGCTATATCTGCGCAACCTGCACTTCTGCAAGTGCAGAGATGAGACTATTC 1270
1271 AAGAAACGAGATCAGCAGAGAGCTTAATGAGACCTTGAGATGCTGATGGGACCGATTG 1330
1272 CAAGACATAGGAGACATGAGACAGAAAGTCCGAGAGAACCGAGGCAATAAGGTTGAT 1331
1331 TACAAGTATAGGAACTATGACAGAGAGCTAGCGAAGACCAAGGATTAAGGTTGAAT 1390
1332 TGAAGAGGAGAGAAACCCAGGCGGCAAGAAATAAATCAAGATATTCACGCTGTGTA 1391
1391 TGAAGAAAGCTGCAAAATCCAAAGTGGCAAGAGAACTCAAGATCTTCACGCTGTGTA 1450
1392 GGCCTCGTGTCTCTTAATGATGATGCTGGGTTCTCAAGTGTGACACAGCTGACTTC 1451
1451 GGCGCTGTGTCTCAAAATGATTTGGCCCGGGTCTGTCACTGGCCAGCCGACTTC 1510
1452 TGTGATTCAGTATGATGATGATTTCTCAACACGAGAGCTTACATGAGATTTCTAAG 1511
1511 GGTGTATCTCAGTATGATGATGATTTCTCAACACGCGGAGAGGACATGATGATTTCTAAG 1570
1512 TTCAAGTAAAGAAACAAACCAAAACCCAAAGGAAAGTCAAGACGAGCAAGCAAAAGTT 1571
1571 CTGAGTAAAGAAACAGAAACCCAAAGCTTAAGAGAAAGATGAAAGTAAACCGAGAGAGCC 1630
1572 CAGTCTTCCAAATGACATGTTCAAGTGGGATTAATAATCTTCTGTGCACAAAGAGCT 1631

```

AUTHORS      Strassberg, R.
TITLE        Direct Submission
JOURNAL      Submitted (17-SEP-2001) National Institutes of Health, Mammalian
              Gene Collection (MGC), Cancer Genomics Office, National Cancer
              Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
              USA
REMARK       NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
              Email: egads@remail.nih.gov
              Tissue Procurement: ATCC
              DNA Library Preparation: Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Genome Sequence Centre,
              BC Cancer Agency, Vancouver, BC, Canada
              info@gscc.bc.ca
              Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
              Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
              Leticia Hsiao, Martin Krzywnski, Reta Kutsche, Oliver Lee, Soo
              Sen Lee, Victor Ling, Carrie Mathewson, Candice McAleavy, Steven
              Nees, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
              Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Scott,
              Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
              George Yang, Scott Zuyderduyn, Marco Marra.

FEATURES
SOURCE
location/Qualifiers
1. 2725
   /organism="Homo sapiens"
   /mol_type="mRNA"
   /db_xref="taxon:9606"
   /clone="MGC:23216 IMAGE:4901057"
   /tissue_type="Muscle, rhabdomyosarcoma"
   /clone_id="MIM_MGC_17"
   /lab_host="DH10B-R"
   /note="Vector: POTB7"
1. 2725
   /gene="DARF1"
   /note="synonyms: DIOL, FL11265, KIA0333"
   /db_xref="LocusID:11083"
   /db_xref="MIM:604140"
238. 1926
   /codon_start=1
   /product="death associated transcription factor 1, isoform
a"
   /protein_id="AAH14489.1"
   /db_xref="GI:15680267"
   /db_xref="LocusID:11083"
   /translation="MDKQDPSBSEAPKAIKPTSKERFKTWGFRFTTIARBGAGDAE
ADLDEPPPOOGLSLRSGRQPRKTRERVSQSLTARRGRSGSPVSLDSGPGC
PATDARTASRCSYBSASRTSGQASSTAVKESAPSKKCGDNDHDSDSNGLT
LKLQNLRLRRKQSEPTERPLKGIQSRLRRKRBBSGPAETVGSBASVTGCVLPSKBE
PNDQGVASQDRCDEWREHSGDVGISBAGRLLENGSDYICPNCTILQVDETHSETAD
HNHRFMICDQAGCEDEPRDICTIGTIRKSSBDGIGRIEIKANPSSKKGLKIFQVYIETAD
QOAKRPRGADADGCTSIGTIRKSSBDGIGRIEIKANPSSKKGLKIFQVYIETAD
GAKICRGCCGCHVAQDSYVCSNDIKLKAATAATWMTFSSGKBEKPKRKMKPKERK
SLPKCAQAGIKTISVHKRPRAPRKETTYKAVVVPARSAALKEAECSTSPWAD
HNHNAVPERKTAAPSRLLYKCYHILGVLGLDPSRSRWIIPWACPELGVALLC"
1048. 1197
   /note="PHD; Region: PHD-finger. PHD folds into an
interleaved type of Zn-finger chelating 2 Zn ions in a
similar manner to that of the RING and FYVE domains"
   /db_xref="CDI:pfam00628"

ORIGIN
Query Match      35.4%; Score 1013.6; DB 9; Length 2725;
Best Local Similarity 73.5%; Pred. No. 4,5e-300;
Matches 1414; Conservative 0; Mismatches 474; Indels 36; Gaps 8;

```

QY	141	GCTTGGCCCTGTGCTTGAACCTGGCCCCAGGCTCTAATATTTATGTAGAAATTTCCAGCCAA	200
Db	148	GCTTGGACCCAGAGGTTTCGTCTGGCAGAGGTTTGTGTGATTTAGAGATTTT-CAGGGAA	206
QY	201	AGTTTCCAAAGCTTTCAGTTTGGGACAGGTATGTATATAAAGGACCTGAGCAATGA	260
Db	207	AAATTCCAAAGCTTTCAGTTTGGGACAGGTATGTATCAAAAGGCGACCCGAGCAATGA	266
QY	261	GAAGACATCCAAAGGCTATCAAAACCAACAGTAAAGAGTTTCAGAAAACCTGGGTTTTCG	320
Db	267	GGAGGCACTTAAGGCATCAAAACCAACAGTAAGATTCAGAAAACATGGGGTTTTCG	326
QY	321	AAGAACCAAGATTGCCAAACGTGAAGGTTGCAGAGACACAGAGGCGACCCCACTGAGCA	380
Db	327	AAAGCACTATTCGCGCAAGCGAGAGGGGCGCAGGGGACCGGAGGCTGACCCACTGGAGCC	386
QY	381	GCAACCC-----ACAGACAGATTAACCTCTCCCTGGCGCGCAGTGGACGGCAACCAAAAG	434
Db	387	GCCACCCCCACAGACAGACAGCTGGGGCTTCTCTGGCGGCGAGTGGAGAGGCAACCCAGCG	446
QY	435	TACTGAGAGGGGTAGAAGATTCTTACACAGTTTCGGCGCGAGGAGAAAAGAAATGTGTC	494
Db	447	CACCTAAGGCGGTGAGAGCATTCCTGACCATTTGCGGGGCGCGCGACAGAGAGCATGCTC	506
QY	495	GGTGTCCCTGAGAGATTCCAGTGAAGCCCACTTTCACAGTCACTGATGTGTGAGACAGC	554
Db	507	TGTCCTCCCTGGAGGATTTCTGTGTAGACCCACGCTCTGGCCCGCACAGACGCGGAGACAGC	566
QY	555	TTCCGAGGGGAGCGTTGAAGACAGTTCTGAGATTCAGAAAGTGGCCCTGATCTGACCTCTT	614
Db	567	CTCCGAGGGCAGCGTGAAAGCGCTTCTGAGACCAAGAGCGGCCCCCACTGTGCTTCCAC	626
QY	615	A---GGGAAAGAACATCTCTGCTCTCTTCTGAAAAGCAAAAGAGGTGAAGAGAAAGAA	671
Db	627	AGCTGTGAAGGAAACGACACAGCCTCTTCTGAAAAGGTGAAGAGGAGATGACACAGATGA	686
QY	672	CACCTCTGACAGTGAACAGTATGAGCTTACGTTGAAGAACTTCAGAAACCGCTTTCGAG	731
Db	687	CACCTCCCAATAGTGAACAGATGAGGCTGACCTTGAAGAAAGCTTCAGAAATCGCTTTCGAG	746
QY	732	AAAGGAGAGCAAGAACCTGTGGAAGAGGTCCTGTGAGGCGAGTCAAGATTCGTCGAGAA	791
Db	747	GAAAGGAGAAACAGAGACCCACTGAAGAGGCCCTGAAAGGAGATCCAGAGTTCGCTTCGAGAA	806
QY	792	GAAAGCAGAGAGGAAAGATTCTGCCGAACCTGAGGTGTCGAATATGAGCAAGTGCAGACA	851
Db	807	GAAAGCGCCGGAGAGAGCGTCCGCCGAGACTGTGGGCTCCAGAGCGCAATGACACATGTGGA	866
QY	852	GGAACAGACCTCTCTGTAAAGACAGAGCCTGAGGCTAGTCAAGGACCAAGTGTCCCACTGAA	911
Db	867	GGGCGCTCTGCGCCAGTAAAGACAGGAAACCCGAGAACATAGAGGGGTTGTGTCTCCAGAGCTGG	926
QY	912	GACACATGACATATGAATAATCAATTTGGAAGGGAAGGCGACTGACGGGAATATCAAGAGGAAA	971
Db	927	GAAACATGACACAGAGACTAAGTTGTAGAGGAAAGGCGCTCAGGACATATCAAGATGAGAGA	986
QY	972	CCCCAGGGAAGGGGCAAAACAAAGCCTGAGTGTAGGTTTACGACCCCAATGACCTGTAA	1031
Db	987	GCGTGGAGACTTGGGGCCGACCGAAGCCTGAAATGTAGGGTTACGACCCCAACGCCCTGTA	1046
QY	1032	CTGACATCTGCGCGCAGCCTCAACAACAAGTTTATGATCTGCTGTGATTCGGTGTGAGAA	1091
Db	1047	TTGCACTTGGCGCGACGCTCAACAACAAGGTTTATGATTTGCTGTGACCGCTGTGAGAA	1106
QY	1092	GTTGTTTCCATGTGACTGTGTGGGTATTTCTGAAAGCCGAGAGGCGCGCTTCGTGAAAGGAA	1155
Db	1107	ATGGTTTCAATGCGATGTGTGTGGGCAATTTCTGAGGCTCGAGGGAGGCTTTTGAAGAGAA	1166
QY	1152	CGGGGAAAGCTACATCTGCGCCAAATTTGACCAATTTGCAAGTGCAGATGTAGACAAACGG	1211
Db	1167	TGGGAAAGCTATATCTGCGCCAAACCTGACCAATTTCTGCAAGTGCAGATGTAGACTATTC	1222
QY	1212	TAGCCCAACCAATGAGCAGAGACTCTGGGTGCAGATCTGTGGTGTCTGATGCGACACAGCTG	1271

Db	1227	AGAAACCGCAGATCAGACGAGACTAAATGGAGACTCGAGAGATGCTATGGCAACGATTTG	1286
Oy	1272	CACAAAGCATNAGGACAGTATGAGCAGAAAGTCCGAGANAGCCAGGGCATTAAGGTTAGAT	1331
Db	1287	TACAAGTATAGGAACAATTAGAGCAGAAAGTCTAGCGAAGAACCAAGGGATTAAGGTTAAAT	1346
Oy	1332	TGAAAGCGACGAACCCGAGCGGCAAGAAAACCTCAAGTATTTCCAGGCTGTGGTAGA	1391
Db	1347	TGAAGAAAGCTGCAAATCCAGTGGCAGAAAGAACTCAAGATCTTCCAGGCTGTGATAGA	1406
Oy	1392	GGCTCTGTGCTCTCTAAATGCAATGGCCCTGTGGTGTTCAGTGTAGCAGAGCTGACTC	1451
Db	1407	GGCCCTGTGTCTCCAAATATATTTGGCCCTCCGGGTGTGTCAAGTGGGAGCCGACTC	1466
Oy	1452	TGTGATTGCAATATGACTGCAATCTCAACACGACAGCACTACATGAGATTTCTTAG	1511
Db	1467	GGTGACTGCAATATGACTGTATCTCAACACGCGCAGCGCAATGAAATTTCTTAG	1526
Oy	1512	TTACAGTAAAGAAACAAAACCAAAACCAAGGAAAAGTCAAGACGAAGCAGAAAATT	1571
Db	1527	CTCAGGTAAAGAAACAGAGCCAAAGCCTTAAGAAAATGAGATGAGACGAGAAAGCC	1586
Oy	1572	CAGTCTTCCAAATATGCAAGTGTTCAGTGGGGATTAATCTCTCTGTGCAAGAGACT	1631
Db	1587	CAGTCTTCCAAATATGCGGTCTCAGGACGATTTAAATCTCTCTGTGCAAGAGAGCC	1646
Oy	1632	AGCGTCAGAGAAAAGGAAACCCAGTGAAGAA---AGTATGCTGCTTCAGAGTGA	1688
Db	1647	AGCTCCAGAAAAGAAAGAGACCAAGTGAAGAGGAGTGTCTCTGCGCGAGTGA	1706
Oy	1689	GACTTTCGGAGAAAGGACGCTGTGAGACGACACCACTCCGGGGCAAGTGCACAA	1748
Db	1707	AGCACTCGGAGAAAGAACCACTGTGTGAGACGACACCCGCTGTGGGAGACCAATCMA	1766
Oy	1749	CTACAAATGCTGTGAAGCAGAGAAAGCAGAGAACCCACTGCACTCTCGCCACCTATT	1808
Db	1767	TTACAAATGCAATTAAGCCAGAAAAGACTG-----CTGCTCCCTGCGGTCAGTGT	1817
Oy	1809	GAGTAAATGTACGTATCAACCAAAAGCTGGCTTCCAGGCGCCCTCCCATATCT-----G	1863
Db	1818	GTATTAATGTATGTATCACTTAGGGGTGTGGCTCTGTGAGACCCCTCCCGATCTTCTGAT	1877
Oy	1864	GCTGCGCTGCTGGGGCTGTCTAGAACAG-----AGTCCGGGGTTCGTGTGCTGATAG	1918
Db	1878	AGCCATCCCTCGGGCTGTCTCAGGACTGGAGATTGCAAGCTTTGTATAGCTGATCAG	1937
Oy	1919	TAGCCAGACAGC---TCACTGCGACGACGAGAGCATACCAAGATGCTCTGACCCAG	1974
Db	1938	ACATCCGCTGACCATCAGCGGAGAGAGAGCCCATATCAGAGATGCTCTGCTGCGCT	1997
Oy	1975	GTTGTTCTGCTAGACTGTGTGAGCTCTCTGTGGGTGTTCTTAAAGCTGTGTAGGCTC	2034
Db	1998	GTTGTCATCTCCTAGTCTGTGAGGACTTCTGTGCACTGTTTCCAAAGCTGTAACTCAC	2057
Oy	2035	ATCT 2038	
Db	2058	TGCT 2061	

RESULT 10					
AX878224					
LOCUS	AX878224	2772 bp	DNA	linear	PAT 17-DEC-2003
DEFINITION	Sequence 13129 from Patent EP1074617.				
ACCESSION	AX878224				
VERSION	AX878224.1	GI:40032960			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE	1				
AUTHORS	Ota,T., Iwagai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,,				

Isihii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
 Primers for synthesizing full-length cDNA and their use
 JOURNAL
 Patent: JP 1074617-A 13129 07-FEB-2001;
 Research Association for Biotechnology (JP)

FEATURES
 source
 Location/Qualifiers

1..2772
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 301..1989
 /note="unassigned protein product"
 /codon_start=1
 /protein_id="CA90061.1"
 /db_xref="GI:40032961"

CDS

ADPLEPPPOO/ALSLRSGKOPKTERVBOPLTARBRGRSPVSLDEGDPSC
 PATDERTASBESVESASSTRSGPQASATVAKRPASSEVKGDDHDTSDSDGLT
 LKEI/LNLRKRKEDEPTERPLKGIQSLRKRREBGPATVSEADIVGVLPSKOB
 PENIDGVVSOAGKDDRESLKBKAODIDEEPDLGRPKPEKIDYVADLCICROP
 HNNRMICCDREBWHGDCVGISSEARGSLERNEDYICPNTCIIQVODTHSETAD
 QOAEWRPDADLDCTDCTSGTIEOKSSDDQIKRIIRKANSQKPKIKIPOPVEAP
 GASKTIGCCCTVAPDPSVYGCNDCLIKHAAATKATLSGSKQKPKPKMKRKP
 SLPKGAAGIKISSVHKRPAPKKRTVKA VVPAPSEALGKELACSSSTPMSAD
 HNYNAVKERTAAPSPDLKYCKMYHGVGLDPSRSFWTAIPWAGPLGVALC"

ORIGIN

Query Match 35.3%; Score 1013; DB 6; Length 2772;

Best Local Similarity 73.4%; Pred. No. 7e-300; Mismatches 475; Indels 36; Gaps 8;
 Matches 1413; Conservative 0; Mismatch 475; Indels 36; Gaps 8;

141 GCTTGACGCTGCTGACCTGACCTGACCTGATTAATTATGATGATTTCCAGCCAA 200
 211 GCTTGACGCTGACCTGACCTGACCTGATTAATTATGATGATTTCCAGCCAA 269
 201 AGTTTCCAGCTTTGATTTGAGGACAGTATGATTAAGGACCTGACGATTA 260
 270 AGTGTCCAGCTTTGATTTGAGGACAGTATGATTAAGGACCTGACGATTA 329
 261 GGAAGACCCCAAGGCTTAATTAATTAAGGATTAAGGATTAAGGATTAAGG 320
 330 GGAAGACCCCAAGGCTTAATTAATTAAGGATTAAGGATTAAGGATTAAGG 389
 321 AAGAACAAGATTTGACCAAGGATTAAGGATTAAGGATTAAGGATTAAGG 380
 390 AAGAACAAGATTTGACCAAGGATTAAGGATTAAGGATTAAGGATTAAGG 449
 381 GGAACCTTTTACCAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTA 434
 450 GCAACCTTTTACCAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTA 509
 435 TACTGAGAGGATTAAGGATTTTCTTACCAAGGATTTGAGGATTTGAGG 494
 510 CACTGAGAGGATTTGAGGATTTTCTTACCAAGGATTTGAGGATTTGAGG 569
 495 GGTGTCCTGAGGATTTGAGGATTTTCTTACCAAGGATTTGAGGATTTGAGG 554
 570 TGTCTCTGAGGATTTGAGGATTTTCTTACCAAGGATTTGAGGATTTGAGG 629
 555 TTCCGAGGAGGATTTGAGGATTTTCTTACCAAGGATTTGAGGATTTGAGG 614
 630 CTCGAGGAGGATTTGAGGATTTTCTTACCAAGGATTTGAGGATTTGAGG 689
 615 A---GGGAAGAACTCTCTCTTCTTAAAGGAAAGGAGGAGGAGGAGGAA 671
 690 ACTGTGAAGAACTCTCTCTTCTTAAAGGAGGAGGAGGAGGAGGAGGAA 749
 672 CACTCTGACAGTATGAGGATTTGAGGATTTTCTTACCAAGGATTTGAGG 721
 750 CACTCTGACAGTATGAGGATTTGAGGATTTTCTTACCAAGGATTTGAGG 809
 732 AAGCGAGAGCAAGAACTCTGAGGATTTGAGGATTTTCTTACCAAGGATTT 791
 810 GAGCGGAGAGCAAGAACTCTGAGGATTTGAGGATTTTCTTACCAAGGATTT 869

792 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 851
 870 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 929
 852 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 911
 930 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 989
 912 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 971
 990 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1049
 972 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1031
 1050 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1109
 1032 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1091
 1110 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1169
 1092 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1151
 1170 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1229
 1152 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1211
 1230 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1289
 1212 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1271
 1290 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1349
 1272 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1331
 1350 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1409
 1332 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1391
 1410 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1469
 1392 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1451
 1470 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1529
 1452 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1511
 1530 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1589
 1512 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1571
 1590 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1649
 1572 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1631
 1650 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1709
 1632 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1688
 1710 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1769
 1689 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1748
 1770 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1829
 1749 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1808
 1830 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1880
 1809 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1863
 1881 GAGCGGAGAGGAGGAGATTTGAGGATTTTCTTACCAAGGATTTGAGG 1940

Db 1470 GCGCCCTGCTGCTCAAAATGATGCGCCCGGCTGCTCACTGCGCGCAGCCGACCTC 1529
Oy 1452 TGTGATGTCAGTAATGATGCTGATCTTCAAAACAGCAGCAGCTAACATGATTTCTAAG 1511
Db 1530 GGTGATGTCAGTAATGATGCTGATCTTCAAAACAGCAGCAGCTAACATGATTTCTAAG 1589
Oy 1512 TTCAGTAAAGAACAAAACAAAACCCAGAGAAAGTCAAGACAGACCGAAGAAATT 1571
Db 1590 CTCAGTAAAGAACAAAACCCAGAGAAAGTCAAGACAGACCGAAGAAATT 1649
Oy 1572 CAGTCTTCAAAATGATGCTGATGCGCGCTTAAATCTTCTGTCAGACAGAGACT 1631
Db 1650 CAGTCTTCAAAATGATGCTGATGCGCGCTTAAATCTTCTGTCAGACAGAGACT 1709
Oy 1632 AGCGTCAGAGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 1688
Db 1710 AGCTTCAGAGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 1769
Oy 1689 GACTTCTGAGAAAGGAGCAGCTGTGAGAGCAGACACATCTTGGGAGAGTACCAAA 1748
Db 1770 AGCACTCGGAAAGGAAAGGAGCTTGTGAGAGCAGACACATCTTGGGAGAGTACCAAA 1829
Oy 1749 CTACAAATCTGAGAAAGGAGCAGCTGTGAGAGCAGACACATCTTGGGAGAGTACCAAA 1808
Db 1830 TTACAAATCTGAGAAAGGAGCAGCTGTGAGAGCAGACACATCTTGGGAGAGTACCAAA 1880
Oy 1809 GAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1863
Db 1881 GTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1940
Oy 1864 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1918
Db 1941 AGCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2000
Oy 1919 TAGCCAGCAGC-----TCACTGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1974
Db 2001 ACACCGGCTGACCATCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2060
Oy 1975 GTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2034
Db 2061 GTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2120
Oy 2035 ATGT 2038
Db 2121 TGTG 2124

RESULT 12
AK002127 2772 bp mRNA linear PRI 01-AUG-2002
LOCUS Homo sapiens cDNA FLJ11265 f18, clone PLACB1009158.
DEFINITION AK002127
ACCESSION AK002127
VERSION 1
KEYWORDS Oligo capping; f18 (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1
Isogai,T., Oca,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiraishi,A., Sudo,H.,
Wagatsuma,M., Hosokiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takehashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,T., Nagahata,K., Masuno,Y., Nimomiya,K. and Iwayanagi,T.
NEO human cDNA sequencing project
Unpublished
JOURNAL 2 (bases 1 to 2772)
REFERENCE Isogai,T. and Otsuki,T.
AUTHORS Direct Submission
TITLES Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

COMMENT (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing; Research Association for Biotechnology: cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
Location/Qualifiers
FEATURES
source
1..2772
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1009158"
/tissue_type="placenta"
/clone_1b="PLACB1"
/note="Cloning vector: pME18sFL3"
301..1989
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA92094.1"
/db_xref="GI:7023815"
/translation="MDKGDPSNEBAPKAIKPTSKFRKTMGRRTTIARBGADAE
ADLPAPPQQOIGSLRSGROPKRTBVBQPLTARBRGRSMPSVLSIDGEPSC
PATDMEIATSGVESASTRSRPGASATVVERPAPSEKVGSDDDTSDPSDEL
LKEIQNRLRRKREBPTEKPLKIQSLRKKRREBGPATVSSASDVTGVLPSKE
PNDQGVSONKDRBSRLBGAADIDEBEPBIDGKPRCEBGPALVYICQIP
HNHRPMICDREWFHEDCVGISSEARGLILRNEDYICPNCTIIQVDETSFAD
OQEARWRGPDADGTCSTIGTIBOKSSBQGLKGRTERKANSKKKLIKIFQVIEAP
GASRIGPCCCHADPVSVCNDCLIKHAAITMTSLSGSKOKPEKRMKPKXP
SLPKCAQAGIKIISVHKRPAPRKETTYKAVVVPABALAKBAKCSSPTSPVMSD
HNYNAVBEKTAAPSPSLYKCMYHGLVGLDLPSSBFWIATIPACPGILVALC"

ORIGIN
Query Match 35.3%; Score 1013; DB 9; Length 2772;
Best Local Similarity 73.4%; Pred. No. 7e-300;
Matches 1413; Conservative 0; Mismatches 475; Indels 36; Gaps 8;
Oy 141 GCTTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 200
Db 211 GCTTGACCAAGGTTTGTCTGCGAGGTTTGTGCTGTTGATTAAGATTT-CAAGAA 269
Oy 201 AGGTTTCAAGCTTTCAAGTTTGGACAGATATGATATTAAGGCACTTACAGCATGA 260
Db 270 AAGTGTCCAAAGCTTTCAAGTGTGGACAGATATGACGACAAAGGAGCAGCCAGCATGA 329
Oy 261 GGAAGCACTCAAGGCTATCAAAACCAACAGTAAGAGTTCAAGAAACCTGGGGTTTTCG 320
Db 330 GGAAGCACTTAAGGCTATCAAAACCAACAGTAAGAGTTCAAGAAACCTGGGGTTTTCG 389
Oy 321 AAGAAACAGATTTGCAAAACGTAGAGTGCAGAGACACAGAGGCGGACCCAGTAGACA 380
Db 390 AAGACACATATCCGCAAGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 449
Oy 381 GCAACC-----ACAGCAGCATTAACCTCTCTGCGCGCGCAGTGTGACCGCAACAAAG 434
Db 450 GCCACCCCAACAGCAGCTGAGGCTGCTGCGCGCGCAGTGTGAGGAGGAGGAGGAGGAG 509
Oy 435 TACTGAGAGGATGAAGAGTTTCTTAACCGGTTGGGCGCGAGGAAAGAAAGATGTGCC 494
Db 510 CACTAGGCGCGTGGAGCAGTTCTTAACCGGTTGGGCGCGCGCGCGCGAGAGAGCATGTCC 569
Oy 495 GGTGTCCCTGAGAGATTCAGTAGAGCCCATCTTCAACAGTCACTGATGTGAGAGCAGC 554
Db 570 TGTCTCTCCGAGAGATTCGTGTGAGCCACGTCCTGCCCGCAGACAGAGCGAGAGCAG 629
Oy 555 TTCGAGGAGAGCGTTGAAGAGATTTGAGATTCAGAAATGAGCGCTGTATTCGATCTCTT 614
Db 630 CTCGAGGAGCAGCGTGAAGAGCGCTTCTGAGACAGAAAGCGCGCCCACTGTCTTCCAC 689
Oy 615 A---GGAAAGAAACATCCCTGCTTCTGAAAGCAAAAGAGAGTGAAGAGAGAGAG 671
Db 690 AGCTGTAAGAGAAACAGCGCTTCTGAAAGAGTGAAGAGAGAGAGAGAGAGAGAGAG 749

CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LML)
http://www.systembiology.org
contact: amadan@systembiology.org

Amup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/LML at: <http://Image.lml.gov>
Series: IRL Plate: 7 Row: 1 Column: 2
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

FEATURES

source

1. 2407
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:3257 IMAGB:3506207"
/issue_type="Kidney, renal cell adenocarcinoma"
/clone_lib="NIH_MGC_14"
/lab_host="DH10B-R"
/note="Vector: pOT87"
1. 2407
/gene="DATP1"
/note="Synonyms: D101, FLJ11265, KIAA0333"
/db_xref="locusid:11083"
/db_xref="MIM:604140"
201. 1835
/codon_start=1
/product="death associated transcription factor 1, isoform b"
/protein_id="AAH00770.1"
/db_xref="GI:12653953"
/xref="locusid:11083"
/translation="MDKSDPSNVEAPKAIKPTSKFRKRWGRRTTIAREGAGDAE
APLPPPOOLGLSLRSGOPKRTVEBOFLTAARRGSMVSLSDSEPTSC
PATDAPTAESGVSASSTRSPQASATVPRPASREKVGCDHDPDSDSDLT
LRLQRLRRKKEQETERPLKGIQRLKRRREBPATVSGEADTYBGVLPSCOE
PNDQVVSQAGQDRESRLRGAQDIDBEPGDGRPRCEGVDPAALYCTCRQP
HNNRPMICDRCEBPHGDCVGISBARGLLERNGDYICPNCTIIQVDETHSERAD
QDAKVRPGDADQCTSIGTIEKSSBQGIKRIEKAAPSGKKLKIPOVIRAP
GASKTGGCCVADDSYVSCNDCLKHAATMKRILSGKXKPKPKMKMKPKPK
SLPKCAQGIKISVHPRPAPRKETTYKAAVVPVARSBALRKAACSSSTPSMSAD
NHNVAVERKTAAPSPSLYKSGKLYSLHPSLIA"

CDS

b

misc_feature
1011. 1160
/note="PHD; Region: PHD-finger. PHD folds into an
interleaved type of Zn-finger chelating 2 Zn ions in a
similar manner to that of the RING and FYVE domains"
/db_xref="CD:pfam00628"

ORIGIN

Query Match 34.4%; Score 985; DB 9; Length 2407;

Best Local Similarity 76.3%; Pred. No. 2.9e-291;

Matches 1266; Conservative 0; Mismatches 380; Indels 13; Gaps 4;

Db 141 GCTGGCCGTCTGCTTACCTGAGCCCAAGCTCTAATTTTATGATGATTCACGCA 200
111 GCTTGACCCAGAGGTTTCTGCTGACGAGGTTTGTGTATTTAGGATTT-CAGGAA 169
Qy 201 AGTTTCAAGCTTCAAGTTTGGACAGGTATGATTAAGAGGACCTGAGCAATA 260
Db 170 AGTGTCAAGCTTCAAGTTTGGACAGGTATGATTAAGAGGACCTGAGCAATA 229
Qy 261 GGAAGCAACCAAGCTATCAACCAACAGTAAGATTCAAGAAAACCTGAGGTTTTCG 320
Db 230 GAGAGCACTTAAGGACATCAACCAACAGCAAGATTCAAGAAAACATGGGGTTTTCG 289
Qy 321 AAGAACCAAGATTGCAACAGTGAAGGTCAGAGACAGAGGCGAGCCCACTGAGACA 380
Db 290 AAGGACCACTATCGCCAGAGAGGCGCAGAGGAGCGGAGGCTGACCACTGAGAGCC 349

Qy 381 GCAACC-----ACAGCAGCATTAACCTCTCCCTGCGCCGAGTGAAGGCAACCAACG 434
Db 350 GCCACCCCAACAGCAGCAGCAGCTGAGGCTCTGCTCCGCGCGAGTGGAGGAGCCCAAGCG 409
Qy 435 TACTGAGAGGTGAGAGATTCTTACACAGGTTCCGCGCCGAGGAGAAAAGATGTC 494
Db 410 CACTGAGCGGTGAGCAGTCTTCTGACCATTTGGCGCGCGCGCGAGGAGGACATGCC 469
Qy 495 GGTGTCCTGAGAGATTCCAGTGAAGCCCATCTCCACAGTCACTGATGTGAGACAGC 554
Db 470 TGTCTCCCTGAGAGATTCTGTAAGCCCACTCTGCTCCGCGCACAGACCCGAGAGC 529
Qy 555 TTCGAGGAGGAGGTTGAAAGCAGTTCTGAGATCAGAAAGGCGCTGATCTGACTCTT 614
Db 530 CTCGAGGAGCAGGTGAGAAAGCCCTCTGAGACCAAGAGGCGCCCAAGCTGCTTCAC 589
Qy 615 AGG--GAAAGACATCTCTCTCTTCTGAAAGGCAAAAGAGTGAAGAGAGAGA 671
Db 590 AGCTGTGAAGGAACGACAGCCTCTTCTGAAAAGTGAAGAGGAGATGACCAAGATGA 649
Qy 672 CACTCTGACAGTGAAGTGAAGTGGCTTACGTTGAAAGGACCTTCAAGACCGCTTGGAG 731
Db 650 CACTCTGATGAGAGAGATGAGATGAGCTTGAAGAGGCTTCAAGATGAGCTTCCAG 709
Qy 732 AAGCAGAGCAAGAACTGTGAGAGGTCCTTGAAGGCACTGAGATCGCTTGAAGAA 791
Db 710 GAAGCGGGAACAGAGACCCCACTGAGAGGCGCTTGAAGAGATCAAGATGAGCTTCCG 769
Qy 792 GAAGCGAGAGAGAGAGATTCTCCCAAACTGGAAGTGTCAAAATAGAGTGCAGTCCAG 851
Db 770 GAAGCGCGGAGAGAGAGGTCCTCCCGAGACTGTGGCTCCGAGGCGAGTCACTGTG 829
Qy 852 GGAAGAGCTCTGTGAAGAGAGGCTGAGCTAGTCAAGGACAGTGTCCAGTGA 911
Db 830 GGGCGTCTGCGCCAGTGAAGAGAGGCGCGAAGAGATCAAGAGGATGTGTCCAGGCTGG 889
Qy 912 GACAGATGACATTAAGAAATCAATTGGAAGAGAGGAGCACTCAAGGAAATCAGAGGAAA 971
Db 890 GAAGATGACAGAGAGATGATGAGAGGAAAGCGCGCTCAAGACATCAAGATGAGAGA 949
Qy 972 CCCGAGGAAACCGGAGCAAAAGCCTGATGTGAGATTGAGACCCCAATGCTCTGTA 1031
Db 950 GCTGAGAGATTGGGCGCGAGCCAGAGCTGAAATGAGAGGTTAGACCCCAAGCGCTGTA 1009
Qy 1032 CTGCATCTGCGCCAGCTTCAACCAAGCTTGAATGATCTGCTGATGATGGGTGAGGA 1091
Db 1010 TTGCATTTGCGCCAGCTTCAACCAAGCTTGAATGATTTGCTGACCGCTGTGAAGA 1069
Qy 1092 GTGATTCATGATGATGATGATGATGATTTCTGAGAGCCCGAGGCGGCTCTGAGAGGAA 1151
Db 1070 ATGATTCATGATGATGATGATGATGATTTCTGAGGCTTCAAGGAGAGCTTTTGAAGAG 1129
Qy 1152 CCGGAGAGACATACATCTGCCCAATTTGCAACATTTTGAAGTGAAGATGAGCAAAACG 1211
Db 1130 TGGGAGAGACATATATCTGCCCAATTTGCAACATTTTGAAGTGAAGATGAGCTATTC 1189
Qy 1212 TAGCGCAACCAATGAGCAGAGCTCTGGTGCAGATCTGTGGGTGAGTGGCAAGCTG 1271
Db 1190 AGAAGAGAGAGATCAAGCAGAGCTTAAATGAGACCTGAGAGATGCTGATGAGCAACG 1249
Qy 1272 CACAGCATTAAGGAGCAGTGAAGAGTCCGAGAGAGACAGAGGCAATAAGGATAGAT 1331
Db 1250 TACAGATTAAGAGCAATGAGAGAGAGTCTAGCGAAGACAGAGGATTAAGGATAGAT 1309
Qy 1332 TGAAGAGAGCAACCCAGCGGCAAGAAAATCTAAGATATTCAGGCTGTCTGTAGA 1391
Db 1310 TGAAGAGAGCTGCAAAATCAAGTGGCAAGAGAACTCAAGATCTTCAAGCTGTGTAGA 1369
Qy 1392 GGTCTCTGTGTCTCTTAATGATTTGGGCTGGGTGTTCAGTGTGACAGGCTGATC 1451
Db 1370 GGTCTGTGTCTCTTAATGATTTGGGCTGGGTGTTCAGTGTGACAGGCTGATC 1429
Qy 1452 TGTGTATTCAGATTAAGTCAATCTCAAAACAGCAGAGGATTAACATGATTTCTAAG 1511

*	201236	202338:	contig of 1123 bp in length
*	202339	205438:	gap of unknown length
*	202439	203500:	contig of 1062 bp in length
*	203501	203600:	gap of unknown length
*	203601	204905:	contig of 1305 bp in length
*	204906	205005:	gap of unknown length
*	205006	206192:	contig of 1187 bp in length
*	206193	206292:	gap of unknown length
*	206293	207385:	contig of 1093 bp in length
*	207386	207485:	gap of unknown length
*	207486	208519:	contig of 1034 bp in length
*	208519	208619:	gap of unknown length
*	208620	209758:	contig of 1139 bp in length
*	209759	209858:	gap of unknown length
*	209859	211441:	contig of 1583 bp in length
*	211442	211541:	gap of unknown length
*	211542	212876:	contig of 1335 bp in length
*	212877	212976:	gap of unknown length
*	212977	214256:	contig of 1280 bp in length
*	214257	214356:	gap of unknown length
*	214357	217003:	contig of 2647 bp in length
*	217004	217103:	gap of unknown length
*	217104	218459:	contig of 1356 bp in length
*	218460	218559:	gap of unknown length
*	218560	218788:	contig of 1229 bp in length
*	219789	219888:	gap of unknown length
*	219889	221547:	contig of 1659 bp in length

FEATURES	Location/Qualifiers
source	1. .221547

```
/organism="Rattus norvegicus"  
/mol_type="genomic DNA"  
/db_xref="taxon:10116"  
/clone="CH230-274116"
```

```
misc_feature 1. .1272 /note="wgs end extension  
clone_end:5p6"  
misc_feature complement(4416. .5263)  
/note="clone_boundary
```

```
clone_end:sp6
site:_
end_sequence:BZ266352
192445..193553
/misc_feature
/notes="wgs_contig"
misc_feature
195743..196592
```

end_sequence:BZ266349"

Query Match	29.8%	Score 853.4;	DB 2;	Length 221547;
Best Local Similarity	81.7%	Pred. No. 2.1e-250;		
Matches 979;	Conservative 0;	Mismatches 107;	Indels 112;	Gaps 1;

53 TTAAGAGCAGTCCCACATGTCCTTTGGTTGTTGCGAGCTCCGAATCTTCATTGAT 112

206293 TTAAAGGCGAGTCCCACTGTCCTTGTGTGTCGAAGCTCCGGAATCTTCATTGAT 206315

Y	Y
113	GAACGTGACTGACAA CAGTGGGGGTGGCGCTTCGCTTGA CTTGGCCCAAGTC 172
206353	GAACGTGACTGACAA CAGTGGGGGTGGCGCTTCGCTTGA CTTGGCCCAAGTC 206412

173 TATAATTTTATGAGATTTCAGCCAAAGTTCCAAAGCTTCAGTTTGGAC----- 227

228 ----- 227

206473 TTTTGTTTATGCGAGTCTCCCCCCAAGATGGAACATGCAGTAACCTTGCCATTACTA 206532

[illegible]

Qy	Db
241	206593
AAAGGGCACCCTGACGAATGAGGAAAGCACCAGGCTATCTAAACCCACCAAGTAAAGAGTTC	AAAGGGCACCCTGAGCAATGAGGAGCAACCAAGGCTATCTAAACCCACCAAGTAAAGAGTTC
300	206552

301 AGGAAACCTGGGGTTTCGAGAACCAACGATGCCAAACGTGAGGGTCAGAGACAGC 360
30653 AGGAAACCTGGGGTTTCGAGAACCAACGATGCCAAACGTGAGGGTCAGAGACAGC 206512

361 GAGCGCGACCCCATGTAGCAGCAACCAAGACATAACTCTTCCTGCGCCGCGAGTGGCA 420

Db 206713 GAGGTGACCCCACTGAGCAGCAACACAGCAGATAACTCTCTCCCTGCGCCGAGTGA 206772

421 GCGCAACCAAACTACTGAGAGGTGAAAGATTCTTACCAACGGTTTCGCGCCGAGG 480

QY

DB	206773	CGGACACCAACCTACTGAGAGGGTAGAGAGTTCTTTACACGGTTGGCGCCGAGGG	206832
QY	481	AAAAAGATGTGCCGGTGTCCCTGGAGAGTTCAGTGAAGCCACATCTTCCACAGTCACT	540

Db 206833 AAAAAAAAAATGTCCTCCCTGAGAGATTCCAGTAGGCCCATCTTCCACAGTCACT 206892

DB	Sequence	Accession
07	GATGTGAGACACATCTCCAGGCGGACCGTGGAAACAGTCTGGATCTAGAAAGTGGCCCT	206893
54	GATGTGAGACACATCTCCAGGCGGACCGTGGAAACAGTCTGGATCTAGAAAGTGGCCCT	206892

QY	601	206953
GTATCTACTCCTTAGGGAAAGACATCTGCCTTTGAAAAAGCCAAAGAGGTGAA	660	GTATCTACTCCTTAGGGAAAGACATCTGCCTTTGAAAAAGCCAAAGAGGTGAA
		207012

Qy	661	GAGGAGGAGGACCTCTGACAGTGACAGTATGACCTTACGTTGAAGGAACTTCAGAAC	720
Db	207013	GAGGAGGAGGACCTCTGACAGTGACAGTATGACCTTACGTTGAAGGAACTTCAGAAC	207072

721 GGCCTTCGGGAAGCGAGCAAGAACTGTGAGAGGTCCTGAGAGCACTCAGAAAT 780
 |||||
 207032 |||||
 207033 |||||

DB 207075 CGCCTGAGAGAGAGCGGAGAGAGAGATTTCTCCGAACTGGAGTGTCCAAATAGGC 840

Db	207133	CGCTGAGGAGGAAGCGCAGAGAGGAAAGTTCTCCGAAACTGGAGTGTCCAAATATGCG	207192
Qy	841	AGTCCCGACAGCA CAGACTCTCTGTATGACAGAGACTGAGGGTAGTCAGGGACCAAGT	900

DB	207193	AGTGCAGACAGGACAGACCTCTCTGTAAACAGGAGCCCTGAGGCTAGTCCAGGACCAAGT	207252
OV	901	TCCCAGTCAGAGACAGATGACATAGAAATCACTTGGAAAGGAAAGCCGACTCAGGAAAT	960

207253 TCCGAGTCAGACAGATACATGAAATCACTTGGAAAGGAAGCGACTCAGCGAAAT 207312

Qy	Db
96	207313
ACACAGGAAAACCCGAGGAAAGCCGCAAAACCAAGCTCTGAGTGTGAGGTTACGACCCC	ACAGAGGAAAACCCGAGGAAAGCCGCAAAACCAAGCTCTGAGTGTGAGGTTACGACCCC
1020	207372

[illegible][illegible]

RESULT 15

AC108337	272404 bp	DNA	linear	HTG 21-SEP-2002
LOCUS				
AC108337				
Rattus norvegicus clone CH230-115113, ***			SEQUENCING IN PROGRESS	
DEFINITION				
***, 2 unordered pieces.				

ACCESSION AC108337
VERSION AC108337.4
KEYWORDS GI:23266147
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

SOURCE ORGANISM
Rattus norvegicus (Norway rat),
Rattus norvegicus
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 272404)

Munry,D,Marie, Mecker,M, Lee, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Albrooks,S, Amin,A, Angiano,D, Ayalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Bielawski,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dedetich,D, Delgado,O, Denison,S, Deramo,C, Ding,Y, Dinh,H, Divya,K, Dwyer,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabriel,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gebregeorgis,B, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W, Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hognes,M, Hollins,B, Howells,S, Hulik,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowalski,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorenuhewa,L, Louisedge,H, Lozada,R,J, Lu,X, Ma,J, Maheshwari,M, Mahidartine,M, Mahmoud,M, Malloy,K, Mangum,A, Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E, Mawhinney,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Milosavljevic,A, Miner,G, Minja,E, Montemayor,U, Moore,S, Morgan,M, Morris,K, Morris,S, Mundaasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,S, Norris,S, Nwokilemeh,O, Okwuonu,G, Olarnpungoon,A, Pal,S, Parks,K, Paternacki,S, Paul,H, Perez,A, Perez,L, Pfannkuch,C, Plopper,F, Poindexter,A, Popovic,D, Prims,E, Pu,L, Puzo,M, Qulroz,J, Rachelin,E, Reeves,K, Regier,M,A, Reigh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rives,C, Rodkey,T, Rojas,A, Rose,M, Rosh,R, Ruiz,S,J, Sanders,W, Savery,G, Scherer,S, Scott,G, Shatman,S, Shen,H, Shetty,J, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smaj,D, Sneed,A, Sodergren,E, Song,X,Z, Sorelle,R, Sosa,J, Steimle,M, Strong,N, Sutton,A, Svatek,A, Tabot,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Uemami,K, Valas,R, Vera,V, Villasana,D, Waldron,L, Walker,B, Wang,J, Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Willson,R, Wlezyk,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Zhou,D, von Weinstock,G, and Gibbs,R.A.

TITLE
JOURNAL

2 (bases 1 to 272404)

Worley,K.C.

Direct Submission

Submitted (27-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 272404)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 21, 2002 this sequence version replaced gi:21737668.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Acllas (<http://www.hgsc.bcm.tmc.edu/projects/rt/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GPMX

Center clone name: CH230-115113

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 238513 bases at least Q40

Consensus quality: 241910 bases at least Q30

Consensus quality: 244206 bases at least Q20

Estimated insert size: 262154; sum-of-contigs estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)

* NOTE: This sequence may represent more than one clone.

* NOTE: This is a 'working draft' sequence. It currently

consists of 2 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

* 270622: contig of 270622 bp in length

* 270623 272404: gap of unknown length

* 270723 272404: contig of 1682 bp in length.

----- Location/Qualifiers

1. 272404

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-115113"

1. 1256

/note="wgs end extension"

clone_end:5p6

complement(1814..2663)

/note="clone boundary"

clone_end:5p6

site:BCORI

end_sequence:BH294125"

62809..64046

/note="wgs contig"

267959..268006

/note="clone boundary"

clone_end:77

site:BCORI

end_sequence:BH294091"

ORIGIN

Query Match 26.1%; Score 748.2; DB 2; Length 272404;

Best Local Similarity 92.1%; Pred. No. 5.8e-218;

Matches 789; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

228 AGGTATGATGATTAAGGACCTGAGCAATGAGGACCAAGGCTATCAACCCAC 128338

AGGTATGATGATTAAGGACCTGAGCAATGAGGACCAAGGCTATCAACCCAC 128338

268 CAGTAAAGGTTTCAAGAAAACCTGGGGTTTGAAGAACCAAGATTGCCAAGGAGG 347

CAGTAAAGGTTTCAAGAAAACCTGGGGTTTGAAGAACCAAGATTGCCAAGGAGG 347

128339 CAGTAAAGGTTTCAAGAAAACCTGGGGTTTGAAGAACCAAGATTGCCAAGGAGG 128338

CAGTAAAGGTTTCAAGAAAACCTGGGGTTTGAAGAACCAAGATTGCCAAGGAGG 128338

348 TGCAGAGACACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 407

TGCAGAGACACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 407

128339 TGCAGAGACACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 128458

TGCAGAGACACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 128458

408 GGGCCGAGTGGACGGCAACCAACGTAAGAGGTTCTTAACCGGT 467

GGGCCGAGTGGACGGCAACCAACGTAAGAGGTTCTTAACCGGT 467

128459 GGGCCGAGTGGACGGCAACCAACGTAAGAGGTTCTTAACCGGT 128518

GGGCCGAGTGGACGGCAACCAACGTAAGAGGTTCTTAACCGGT 128518

```
OY      468 TCGCCGCCGAGGAGAAAGAAATGTGCCCGGTGTCCCTGAGAGATTCCAGTGAAGCCCAATC 527
Db      128519 TCGGCGCCGAGGAGAGAAAATGTGCCCGGTGTCCCTGAGAGATTCCAGTGAAGCCCAATC 128578
OY      528 TTCACAGTCACTGATGTGAGACAGCTTCCGAGGGGAGCGTTGAAGCAAGTTCTGAGAT 587
Db      128579 CTCACAGTCACTGATGTGAGACAGCTTCCGAGGGGAGCGTTGAAGCAAGTTCTGAGAT 128638
OY      588 CAGAAAGTGGCCCTGATCTGACTCTTTAGGGAGAAACATCCCTGACCTTTCTGAAGAGGC 647
Db      128639 CAGAAAGTGGCCCTGATCTGACTCTTTAGGGAGAAACATCCCTGACCTTTCTGAAGAGGT 128698
OY      648 AAAAGAGGTGAAGAGAGAGAACACCTCTGACAGTGAACAGTATGACCTTACGTTGAA 707
Db      128699 GAAAGAGGTGAAGAGAGAGAACACCTCTGACAGTGAACAGTATGACCTTACGTTGAA 128758
OY      708 GGAACCTTGAAACCGCCTTCCGAGAAAGCGAGAGCAAGAACCTTGTGAGAGGTCCCTGAG 767
Db      128759 GGAACCTTGAAACCGCCTTCCGAGAAAGCGAGAGCAAGAACCTTGTGAGAGGTCCCTGAG 128818
OY      768 AGGCGATGAGATGGCTGAGAGAGAGGCGCAGAGAGAGATTTCTGCCGAAACTGAGAG 827
Db      128819 AGGCGATGAGATGGCTGAGAGAGAGGCGCGAGAGAGAGATCTGCCGAAAGCTGAGAG 128878
OY      828 TGTCCAATTAGCAGTGCAGAGCAGAGACGACCTCTCTGTAAAGCAGAGCCTGAGGCTAG 887
Db      128879 TGTCCAATTAGCAGTGCAGAGCAGAGACGACCTCTCTGTAAAGCAGAGCCTGAGGCTAG 128938
OY      888 TCAGGAGCCAGTGTCCCACTCAGAGACAGATGACATAGAAATCACTTGGAAAGGAGGC 947
Db      128939 CCAGGAGCCAGTGTCCCACTCAGAGAGAGATGACAGAGAAATCACTTGGAGGAGAGGC 128998
OY      948 GACTCAGGGAATTAAGAGGAGAAACCCGAGGAGAGCGGCAAAACCAAGCTGAGGTGA 1007
Db      128999 GACTCAGGGAATTAAGAGGAGAAACCCGAGGAGAGCGGCAAAACCAAGCTGAGGTGA 129058
OY      1008 GGTTTACGACCCCAATGCGCTGTACTGCACTTGCCGCGCAGCCTCACAAACAGGTTTAT 1067
Db      129059 GGTTTACGACCCCAATGCGCTGTACTGCACTTGCCGCGCAGCCTCACAAACAGGTTAGGT 129118
OY      1068 GATCTGCTGTGATCGGT 1084
Db      129119 GGAGGAGGGAACCTGGT 129135
```

Search completed: April 23, 2004, 14:16:35
Job time : 10938.8 secs